PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] : H04N 7/18 A1

(11) International Publication Number: WO 97/08896

(43) International Publication Date: 6 March 1997 (06.03.97)

(21) International Application Number: PCT/US95/10681

(22) International Filing Date: 23 August 1995 (23.08.95)

(71) Applicant (*for all designated States except US*): SCIENTIFIC-ATLANTA, INC. [US/US]; One Technology Parkway, South, Norcross, GA 30092-2967 (US).

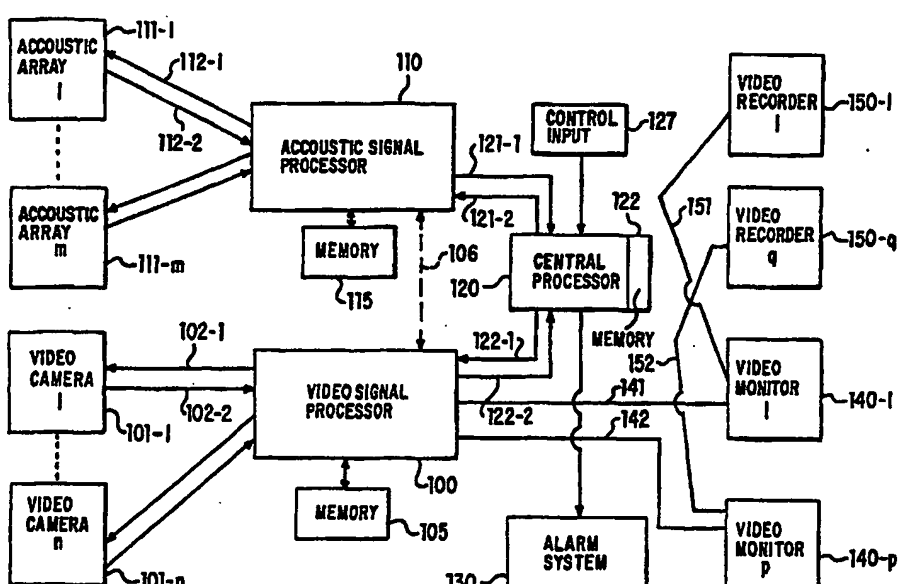
(72) Inventors; and
(75) Inventors/Applicants (*for US only*): RAMSAY, Leigh, William [US/US]; 13821 Frame Road, Poway, CA 92064 (US). DALRYMPLE, Dale, Brian [US/US]; 3322 Cowley Way #3, San Diego, CA 92117 (US).

(74) Agents: POTENZA, Joseph, M. et al.; Banner & Allegretti, Ltd., 11th floor, 1001 G Street N.W., Washington, DC 20001-4597 (US).

(81) Designated States: CA, JP, MX, US.

Published
With international search report.

(54) Title: OPEN AREA SECURITY SYSTEM



(57) Abstract

An open area security system comprises an acoustic sensor array (111-1, ..., 111-m) capable of forming elevational and azimuthal beams, or comprises two such arrays separated by a predetermined distance. A camera (101-1, ..., 101-n) mounted in the vicinity of the arrays may be automatically directed toward a detected, sound-producing event. Event data may be prestored in memory (105) and the system may learn of the event's character as an emergency or non-emergency status. Triangulation and other computational techniques may be utilized to determine from the beams the location (x/y coordinates) of the event, thus allowing the camera to be focused and zoomed to capture high resolution images of the event.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

OPEN AREA SECURITY SYSTEM

BACKGROUND OF THE INVENTION

1. Technical Field

The present invention generally relates to the field of security systems and, more particularly, to the field of audio and video surveillance systems for open areas. The system of the present invention includes acoustic sensor arrays for collecting sound (acoustic) signals relative to activities or events occurring in a particular open area protected by the system, and a data processing and control system for processing the received signals, differentiating among events, classifying the events, for example, as emergency related events, and automatically embarking on or recommending particular courses of action, e.g., pointing a video camera in the direction of a detected event.

2. Description of the Related Arts

Acoustic sensor arrays are known from the art of submarine warfare, for example, which comprise, for example, a plurality of as many as 1200 microphones which are adapted to form elevational beams and azimuthal beams. Of course, a microphone is a transducer which is capable of collecting sound signals. A microphone converts sound signals to electrical signals according to the frequency response of the microphone. If sound is received, for example, via a linear, circular or other array of such microphones, signal processing can be performed to obtain accurate beam formation in a plane of interest. The converted acoustic to electrical signals are delayed and summed together and further processed, for example, via an acoustic signal processor based on the relative time relationship the sound signals are received at the different microphones of the arrays. The detected frequency response of the received signals over time is used to distinguish one sound from another. For example, in submarine warfare, a submarine of a foe may be distinguished from a submarine of a friend; the propeller sounds of a surface travelling tanker ship may be identified. Whale, porpoise and other sounds made by fish may be differentiated from one another. Moreover, the distance to a sound source and direction from which a sound signal is received may be determined.

Many of the principles of sound engineering are described in the textbook Sonar Engineering Handbook, by Harrison T. Loeser, 1992, available through Peninsula Publishing, Los Altos, California. Infrasonic frequencies are described as those below audible sound or at frequencies between, for example, 0 and 20 Hz. Audible sound is characterized as sound between, for example, 20 Hz and 20,000 Hz, while ultrasonic frequencies are those above

20,000 Hz. Two types of sound wave spreading are spherical and cylindrical. Spherical spreading occurs when the sound spreads uniformly over a sphere (or hemisphere) that expands with distance. The transmission loss of a propagating sound wave varies as the inverse of the square of the radius of the sphere. Cylindrical spreading occurs when the sound spreads uniformly over a cylinder that expands with distance. The transmission loss of a propagating sound wave varies as the inverse of the radius of the cylinder. Beamforming is the process of listening to sound from an array at selected elevational and azimuthal angles. It reduces the unwanted noise at a processor by amplifying the signals arriving from the selected angle and provides bearing and depression/elevation angle information concerning the source of the sound. Shading the responses of phones of an array may be used to improve the main lobe of a phone response and reduce the side lobes. Shading refers to increasing or decreasing the gain on a phone signal before it goes to the processor. Side lobe level and beamwidth can be controlled also by varying the spacing of the array elements. Element spacing may be geometrically tapered or otherwise spacially placed. Spacial tapering can permit higher resolution or a significant reduction in the number of elements. The beamformer provides the proper time delays and shading of signals from the phones of the array and sums them to form the input from the selected angle. The signal is then transmitted to the processor.

The processor of acoustic signals (that have been converted to electrical signals) must comprise a variety of algorithmic manipulations of recovered signals. The electrical signals are sampled over time, preferably such that the collected samples exceed twice the bandwidth times the time of data collection. Fast Fourier transform analysis is used to break down the signal into a collection of frequencies having varying amplitudes over time. The collected data is probabilistically correlated for classification purposes and stored for comparison with new sounds. The processing is described to some extent in Loeser identified above and also by A. Winder and Charles J. Loda in their text Space-Time Information Processing, 1981 edition, also available from Peninsula Publishing.

In the art of security systems and surveillance systems, it is known to monitor ambient sounds and identify human or animal cries, for example, from U.S. Patent No.'s 4,131,887; 4,237,449; 4,365,238; and 4,853,674. To assist hearing impaired individuals, for example, according to the '674, '238 and '449 patents, audible signals such as automobile horns, police sirens, crying babies and the like may be differentiated by frequency and amplitude and cause an appropriate indicator to selectively actuate so the hearing impaired individual may be

assisted to an appreciation of the triggering event. The '887 patent suggests that a barking dog sound may be identified and, subsequently trigger, for example, floodlights and/or a remote indicator at a guard or police station.

It is also known to differentiate the sound of broken glass as taught by U.S. Patent No.'s 4,060,803; 4,241,335; 4,668,941; 5,164,703; 5,323,141; 5,376,919;5,414,409; and 5,428,345. According to the '941 patent, the sound of breaking glass comprises a low frequency thump sound followed by a tinkle sound. In other words, the characteristic frequency and amplitude characteristics of a received sound may be identified over time as the sound of broken glass.

It is also known in such arts to actuate something other than an LED indicator or call to a remote police station and to attempt to localize the direction of the sound. For example, U.S. Patent No. 4,806,931 teaches to identify, for example, a police or rescue squad siren and localize the direction from which the sound is coming in order to actuate traffic signals to permit traffic flow only in the direction from which the sound is received. Of course, the advantage of such a system is that traffic accidents with emergency vehicles may be prevented.

More recently, it is known from U.S. Patent 4,920,332 to adapt a threshold level of an aperiodic wave resulting, for example, from a door opening to discriminate between alarm and non-alarm events. Also, according to U.S. Patent 4,935,952, an energy discriminator differentiates a fire-alarm acoustic signal from background noise and triggers digital circuitry for dialing an emergency number and triggering an audio interface. Once the audio message is delivered, the line is automatically hung up to permit a return confirmation call.

Video cameras are known for providing surveillance of an open area. Video cameras, however, alone do not give complete coverage of the area, and observers do not necessarily detect all incidents or events to permit response in a timely manner. Human operators are frequently expected to view large numbers of monitors for long periods of time. Boredom, fatigue, psychological and other effects may prevent operators from identifying emergency activities, classifying them and acting appropriately to, if possible, limit the risk of loss of property or life and assure capture of the event in as efficient and complete a manner as possible. While it is also known to utilize video tape backup and coordinate their use with time-of-day measurements, there is no assurance that the captured event will be captured in sufficient detail, for example, to assist in suspect identification or in prosecution.

Such known surveillance systems, consequently, suffer from their inherent difficulty in adapting other than predetermined threshold levels and/or differentiating among a plurality of emergency and non-emergency events. Moreover, none of the disclosed systems teach or suggest their being coupled to video or other camera surveillance systems to focus received camera images to an area of interest. Moreover, none of the disclosed systems include diagnostic and control systems for differentiating and classifying identified events and embarking on a plurality of different response scenarios depending on the identified event.

Consequently, it is an object of the present invention to provide an improved system for monitoring the security of an open area.

It is a further object of the present invention to apply multiple forms of information gathering devices including, for example, acoustic sensor arrays, still or motion cameras, laser and infrared sources and receivers as well as known human video observation.

Moreover, it is an object of the present invention to apply complex adaptive processing (e.g., artificial intelligence) systems for recording and learning typical and atypical events that may occur in an open area protected by the system.

It is also an object of the present invention to train cameras on activities identified by localization and differentiation of those activities which are not in conformity with learned or typical or atypical, recorded events.

It is also an object of the present invention to utilize video signal processing techniques in real time, for example, to permit a camera to follow moving objects and off line, for example, to obtain suspect identification data or collect criminal evidence.

It is a still further object of the present invention to provide an open area surveillance system for classifying among events as suspicious, hostile and friendly or other classifications and, besides actuating one or more cameras to pan, tilt and/or zoom to an area of interest, to recommend a course of action and/or automatically initiate at least preliminary steps of the recommended course of action.

SUMMARY OF THE INVENTION

The problems and related deficiencies of prior art surveillance and security systems are overcome by the principles of the present invention, an open area security system using one or more arrays of microphones, for example, mounted on poles such as parking lot light poles to receive acoustic signals in horizontal and vertical planes. Moreover, the system includes pole-mounted cameras which may be automatically controlled to pan, tilt, rotate and/or zoom on an identified area of interest. Sounds typically occurring in the open area

to be protected, for example, in a parking lot, of starting automobiles, automobile engines running and the like may be stored in digital form in memory and stored as a library of typical sounds. Similarly a library of atypical sounds may be stored for atypical events such as a woman screaming or a pistol firing. Subsequently, a received sound signal pattern comprising frequency response or other signal characteristics (e.g., cepstral or LPC coefficients) over time plots is compared with stored sound patterns in the library of sound patterns and identified and/or differentiated from other stored sounds. The acoustic array signal input is used to obtain a digital "signature" used herein to signify a reference to a predetermined acoustic pattern that is obtainable from recording an event via an acoustic array and is, consequently, pattern recognizable. Moreover, by operating the array in two planes and, if appropriate, via plural locations (for example, plural light poles) events are automatically ranged as to distance and direction via one or more of the following ranging methods: beam intersection (i.e., triangulation), dual vertical sensor correlation and/or dual horizontal sensor correlation. The individual locations may communicate with a central location by wireless (such as radio frequency) means or cable.

At a central monitoring location, for example, multiple video monitors and video cassette or other recorders may be used to observe and record as in the prior art but, according to the present invention, a diagnostic and control system is provided for learning and differentiating events, controlling camera operation, classifying events and automatically operating according to a recommended course of action depending on the event. For example, a victim is attacked in a monitored open area such as a parking lot, the acoustic conical array by the methods described above localizes the sound and differentiates the sound from normal, typical, prerecorded and digitally stored events and because the present event is classified as a victim's scream, ranges the event as to direction and distance and so actuates a particular local camera to focus on the direction of the sound at a distance determined through the sound differentiation process. Moreover, once the sound is differentiated as a victim's screams, an alarm may be sounded and armed officers dispatched to the scene. Simultaneously, a video processing system may be actuated to process the received video, recognize attacker movement and cause a train of cameras to follow the attacker as the attacker attempts to escape the scene of the attack. Thus, detection of a hostile event, identification of the event, direction of cameras, classification of the event, responsive action other than camera direction and recordation of the event may all be efficiently undertaken according to the principles of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block schematic diagram of one embodiment of a system according to the present invention including an acoustic array for collecting sound signals and providing them via, for example, a wireless link to an acoustic signal processor and a video camera for capturing video images and providing them to a video signal processor. A user operating a control input of a central processor coupling the acoustic and video processor may control either the array or the camera, or the central processor may automatically operate to classify events captured by either and operate on those events by sounding alarms, directing camera movement or taking other appropriate action or making recommendations based on reasonable inferences.

Figure 2 is also a schematic block diagram of a system according to the present invention which eliminates the requirement for a central control processor.

Figures 3a and 3b comprise combination apparatus and flowcharts for describing acoustic signal processing prior to beamforming and determination of bearings and post-detection beamforming and determination of bearings.

Figures 4a and 4b comprise drawings showing the approach of sound waves perpendicular to a linear array (Figure 4a) and the approach of sound waves at an angle such as 45 degrees to the array.

Figures 5a and 5b comprise drawings similar to those of Figures 4a and 4b for a circular array.

Figure 6 shows a circular array of microphones and is useful for describing the concept of equivalent aperture and pseudophones.

Figure 7 is a first figure for explaining the present invention in the context of a particular application such as monitoring an open area such as a parking lot, train track or platform of a mass transit transportation station.

Figure 8 is a second figure for describing a parking lot security system design, the particular design showing an arrangement for an approximately 400 foot by 800 foot parking area.

Figure 9 is a figure showing a sector command and control center and its connection to a central command and control center for managing a parking lot security system involving a plurality of monitored open areas (or sectors).

Figure 10 is in part a chart and in part a flowchart useful for describing the principles of the present invention.

Figures 11-13 comprise amplitude versus frequency (frequency response) over time plots of different acoustic signals: Figure 11 represents a pistol firing, Figure 12, a woman screaming and Figure 13 a train inbound/outbound.

Figure 14a comprises a flow diagram showing typical operations and actions useful for explaining the operation of the present invention in conjunction with a particular event, namely a woman screaming.

Figure 14b shows a typical memory table that may be constructed in memory for associating identified sounds with courses of actions.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring now to Figure 1, there is shown an overall block diagram of an open area security system according to the present invention. Generally, similar reference numerals or characters are used throughout the following description to refer to similar elements. Briefly referring to Figure 7, there is shown a video camera 723 and first and second acoustic arrays 721-1 and 721-2 mounted to a pole 701 (also conveniently used for lighting the open area to be protected). Figure 7 only shows one open area or sector but a plurality of such open areas may be protected in accordance with the principles of the present invention as will be further discussed in connection with Figure 9.

In Figure 1, however, acoustic arrays 721-1 and 721-2 are more generally shown as acoustic arrays 1 . . m and are referenced 111-1 to 111-m where m is the total number of acoustic arrays. In other words, there may be as many acoustic arrays as required for protecting a particular open area or plurality of areas. An acoustic array may be linear or a linear row of microphones, may be circular or may be otherwise shaped for forming beams in different planes of interest according to known principles. Some of the criteria for selecting the number and construction of acoustic arrays to be provided in a system according to the present invention will be developed as the description continues.

Similarly, video camera 723 of Figure 7 is shown generally as video cameras 1 . . n and referenced by reference numerals 101-1 to 101-n, the cameras being used for monitoring events and activities at a protected open area or areas. The cameras are preferably video cameras but may comprise high resolution digital still cameras now known in the art available from Eastman Kodak Company which are actuated at or near video rates of thirty frames per second. The images that are captured are preferably digital and, if RF communications are used to transmit the signal, the signal may be compressed prior to transmission, for example, in accordance with known compression standards such as MPEG 2 video compression

standards. Some of the criteria for selecting the character and number of cameras to be provided for a single area will be developed as the description continues.

To control the arrays and associated circuitry, control leads or paths 112-2 are shown which are preferably not hard wired but represent wireless communication links for controlling the arrays. Output signals, typically collected electrical signals representing collected converted acoustic to electric signals are output via output paths 112-1 which are likewise preferably not hard-wired but comprise a wireless communication path. Path 112-1 and 112-2 couple the acoustic arrays with an acoustic signal processor 110.

To control the cameras and associated circuitry (such as pan zoom control 722 of Figure 7), control leads or paths 102-1 are provided which are preferably not hard wired but represent wireless communication links for controlling the cameras. Output signals, typically compressed video signals are output via output paths 102-2 which are likewise preferably not hard-wired but comprise a wireless communication path. Path 102-1 and 102-2 couple the camera to a video signal processor 100. The acoustic and video paths to processors 100 and 110 and reverse control channels preferably comprise channels of a radio frequency transmission link that may be microwave, UHF, or other frequency and may involve so-called LLEOSAT or other satellite transmission.

Power for each of the cameras and acoustic array circuits and control and communication circuits is conveniently provided by the same power that is provided to a pole-mounted light (not shown) also mounted to a pole 701. Wireless communication from the pole 701 is preferable to save the costs of running additional communications wires or optical fibers, if not already available. If a parking area or other open area is not yet constructed, the link may preferably be a wire, cable or fiber optic link. The cameras may be supplemented via infrared or other invisible light, and these lighting systems similarly powered by the same source.

Acoustic signal processor 110 is preferably a processing algorithm controlled processor which may be a processor operating in parallel with video signal processor 100 or in tandem via separate processor machines. One purpose of acoustic signal processor 110 is to formulate a beam, determine its bearing and process the beam in conjunction with previously stored beam representations (as will be further discussed in conjunction with Figures 11-13) to identify the event, classify the event and initiate a course of action as will be further described herein. For example, one course of action may be to signal the video processor 100 to control the cameras 101-1 to 101-n to follow an ongoing event.

Video signal processor 100 is preferably a processing algorithm controlled processor which may be a processor operating in parallel with audio signal processor 110 or in tandem via separate processor machines. One purpose of video signal processor 100 may be to determine a situs of movement via a movement detection algorithm known in the art and, thus, operate to signal the acoustic signal processor to signal the acoustic arrays to focus their attention on a particular movement in a protected area. Another purpose is to zoom to an event and thus obtain a record via coupled monitors or video recorders of as high a resolution image as may be possible for, for example, subsequent suspect identification purposes. The video processor determines a portion of a captured image in which movement is determined and can zoom the camera to envelop that determined portion from evaluating pixel values that are equal (remain unchanged over time) or are varying in intensity.

Optional bidirectional path 106 is shown here as coupling the audio signal processor 110 and the video signal processor 100 (and in Figure 2, this embodiment is explored further). These processors 100 and 110 may be considered as the same computer workstation or parallel workstations or otherwise designed in accordance with different operating systems according to well known computer systems engineering principles. For example, path 106 may comprise a bus system joining two co-processors or may, as represented in Figure 1 comprise bidirectional paths 121-1 and 121-2 and 122-1 and 122-2 connecting the processors 110 and 100 with a master or control processor 120 having its own memory 122.

Memories 105, 115 and 122 coupled to processors 100, 110 and 120 respectively may, in fact, comprise different sections of the same memory or different memories. Typically, algorithms and other permanently stored data are preserved in read only memory, and the memory is not volatile or destructible on loss of power. Other data is stored temporarily in temporary or random access memory as is well known in the art.

According to the principles of the present invention, each of processors 100 and 110 comprise complex adaptive processing systems which "learn" patterns of customary and emergency events through human intervention. Tables are established in memory 105, 115 or 122 of events that can be identified as they recur. Once the event is stored, the event can be labeled as emergency or non-emergency and a particular automatic course of action established which is automatically initiated or overridden by human intervention.

Via link 131, for example, an alarm system 130 may be triggered which may comprise any known alarm system suitable for the present purpose. For example, alarm system may comprise loudspeakers arranged on poles 701 or remote signal systems for alerting police,

fire and rescue services to dispatch emergency vehicles. Other alarm system arrangements may only be bounded by economic and practical considerations.

Control input 127 may comprise a keyboard, mouse, joystick or other arrangement or combination of arrangements which may be utilized for accomplishing camera control, switching of signals to monitors and recorders, acoustic array control, alarm system control and assisting in complex adaptive processes for processing acoustic and video signals. For example, if a mouse is used, typically central processor 120 will further include a display monitor (not shown) for facilitating each of the above-identified functions and others.

It is assumed that the images of an event may be captured, viewed and permanently recorded. In some respects, it may be most advantageous to store high resolution digital video images including sounds in memory 105 and more routine low resolution video on video tape or disc via recorders 150-1 to 150-q. Certainly, there exists a trade-off in memory costs between high and low resolution storage that can be managed via the complex adaptive processing system of the present invention. For example, routine events may be routinely videotaped while emergency events (which occur relatively quickly) may be stored in memory 105, 115 or 122 as high resolution digital representations (audio or image). Video monitors 140-1 to 140-p may be conveniently utilized for viewing events at a protected open area or plurality of protected areas. These are shown connected directly with video recorders but more than one signal may be passed on channels 141 and 142 and selectively switched via control leads or otherwise addressed and gated to monitors and recorders via leads 151 and 152.

Figure 2 shows a second embodiment of a system according to the present invention. Audio signals and alarms are provided via leads 112-1 (which may be preferably wireless) to acoustic processor 110. Lead 106 is described as audio beam selection and control and, rather than having a central processor, control is distributed to video processor 100 so that acoustic processor 110 operates as a slave thereof. Camera and audio control 104 is shown as a mouse control in this embodiment. Other depicted elements perform similarly as the similarly labeled elements of Figure 1.

Figure 3a and 3b may be viewed together as describing two approaches for the processing of acoustic array data. Depicted are vertical phone arrays which may comprise part of cylindrical arrays as further described herein below. In each approach, the initial beam forming processes, operating on vertical phone arrays, produce elevational beams at varying angles. A first approach involving pre-detection beamforming is shown in Figure 3a.

A second approach involving post-detection beamforming and bearing determination is shown in Figure 3b.

In both Figures 3a and 3b elevational beamforming is performed prior to any other processing. This eliminates contamination from overhead noise sources such as passing airplanes. In Figure 3a azimuthal beamforming is performed prior to acoustic event detection. That is, a series of azimuthal beams a few degrees wide are created and the signal in each beam is scanned for acoustic trigger events. In Figure 3b all the data in a series of azimuthal beams is summed to provide a relatively "flat" beam which is omnidirectional in the horizontal plane. This omnidirectional data is then scanned for acoustic trigger events. If one is detected then azimuthal beamforming is performed in all directions and the event re-detected in a specific beam. Figure 3a requires substantially more processing than Figure 3b, but may be able to reliably detect events at a lower signal-to-noise ratio.

Referring specifically to Figure 3a, there is shown a plurality of linear (which may be vertically arranged) arrays. A first beamforming array, beamform 1 represented as element 305, is shown including microphones 300-1 to 300-n (this n bearing no relationship to the n shown in Figure 1, 2 or the n of "Beamform n" shown as element 306). The number of microphones in a vertical array, for example, may comprise from two to sixteen. An nth beamforming array 306 comprises microphones 301-1 to 301-m (this "m" bearing no relationship to "m" used in Figure 1, 2 or the m shown in circle 310 as "Beamform m beams").

The sound signals impinging on each of arrays Beamform 1 305 and Beamform n 306 are forwarded to step 310 where m beams are formed as will be subsequently described in further detail herein. Briefly, the step of beamforming comprises summing the output of all the phones of the array, utilizing suitable signal delays. A sound wave approaching the array of phones will be amplified according to the sum of the phone outputs, thus producing the so-called gain of the array. For example, a sound wave approaching perpendicular to a linear array will create a gain without signal delaying since the sound wave will be received at all the phones of the array at approximately the same time. However, since there is a predetermined distance between the phones of an array, a sound wave approaching from one side or another will strike one phone after another resulting in less than a perfect summation of the signals and less gain. Electronic signal delay can be used to compensate for the positional differences of the phones. The beamforming process thus can be used to maximize the signal output for sound waves emanating from different directions.

The proposed system may comprise a plurality of circular arrays stacked, for example, eight phones high. If a circular array comprises eight microphones each separated at 45 degrees, and the arrays are stacked eight high to form eight vertical linear arrays, then a total of sixty-four phones will form one cylindrically shaped array capable of forming both elevational and azimuthal beams. Other array shapes are possible than cylindrical shapes, such as spherical or other shapes appropriate for the protected area. The output of the m beam beamforming step is forwarded to processing step 330 and to operator audio selection step 320.

Operator audio selection step 320 receives beam select control inputs and provides beam audio outputs. Beam select control is the operator's request via some input device to receive audio from a specific beam. Beam audio output is the corresponding beam audio, played through an audible acoustic channel, typically speakers or a headphone. The purpose of selection step 320 is to allow an operator to listen to the acoustic signals in a specific beam. Each beam is pointed in a different direction horizontally and vertically, so acoustic events will sound the loudest in beams which point toward the source. Step 320 involves basically an audio channel selection, wherein each audio channel corresponds to a unique direction in the horizontal and vertical planes.

Beam processing step 330 is primarily undertaken via acoustic processor 110. The m beams are processed as follows: filtered to particular band of interest or to eliminate noise, via fast fourier transform operations, via averaging with other inputs, via equalization in the time domain and in the space domain according to the plane of interest, via detecting a certain threshold level of amplitude and via 1 of m beam selection prior to initiating alarms and bearing to acoustic signal processing stage 2.

Now referring to Figure 3b, the same vertical microphone arrays are shown comprising microphones 300-1 to 300-n and 301-1 to 301-m and the same beamforming steps 305 representing beamform 1 and step 306 representing the formation of the nth beam. At step 340, all azimuthal beam data is stored in memory, for example, memory 115. All azimuthal beam data is summed together at step 350 and a step 360, analogous to step 310 of Figure 3a, is performed for the beam in which an event has been detected.

The outputs of steps 340 and 350 are accessible to further processing step 370 and involves further filtering, transformation via fast fourier transform, averaging, noise equalization, temporal equalization, spatial equalization, 1 of m beam selection and, in this

step, a bearing computation for the event. Alarms and bearing computations are forwarded to an acoustic signal processor if processing step 370 is considered a pre-processing step.

In each of steps 330 (Fig. 3a) and 370 (Fig. 3b), a set of acoustic data is scanned for trigger events, such as the sound of a gunshot. In step 330, the data being scanned is limited to a specific angular region both horizontally and vertically. In step 370, the data is limited only vertically, and is omnidirectional in the horizontal plane.

Figures 4a and 4b taken together show how a sound wave may approach a linear array differently. Figure 4a shows a sound wave approaching a linear array comprising microphones 400-1 to 400-n from the right, perpendicular (at 90 degrees) to the array. The signals are summed at step 410 and forwarded as a relatively large amplitude output for further processing because a sound wave strikes all phones of the array at approximately the same time.

In Figure 4b on the other hand, sound waves are shown approaching from a 45 degree angle and, thus, arrive at different times at microphones 400-1 to 400-n. For example, a sound wave approaching from the upper right of the drawing impinges on microphones 400-1, -2 and so on sooner than microphones 400-(n-1), 400-n. Consequently, the summing step 410 results in a relatively small amplitude signal. The gain of the array is consequently less than in Figure 4a. Beam formation through the use of signal delay means can equalize the gain for sound waves emanating from different directions.

Figures 5a and 5b are provided to show typical approaches of sound waves to a circular array, for example, for forming a horizontal beam according to a preferred embodiment of the present invention. Figure 5a shows a relatively perpendicular wave approach. Figure 5b shows an approach of a wave at a 45 degree angle. Both figures show microphones 500-1 to 500-8 forming the array (any number of suitably placed microphones may be used at preferably equivalent known distances from one another). The microphones 500-1 to 500-8 are shown separated at 45 degree increments around the circle. A circular array for a pole 701 will typically have a diameter of about one meter or less. Much larger sizes could become unwieldy. However, the size of the array determines the lowest frequency that can be processed directionally. In general, the array must be at least as wide as the sound wavelength at the frequency of interest. A one meter array should be adequate to approximately 300 Hz.

The spacing and size of phone arrays is inversely proportional to the design frequency response. If one doubles the design frequency, then one doubles the number of phones

required for a given aperture. There exist known digital signal processing techniques to reduce this effect but the proportionality generally still holds. Of course, more phones mean greater costs and an attendant greater requirement for parallel processing and memory access. A band of from 0-3000 Hz generally should suffice for the present security purposes; however, arrays for higher frequency signals may be provided if the need is demonstrated in view of costs.

A higher frequency designed array would have the advantage of sound differentiation at a higher level of resolution. For example, one could recognize a particular type of siren and determine with accuracy its location. Most sirens have fundamental frequencies below 1 kHz and are therefore detectable. However, the relative power in the harmonics (at higher frequencies), if detected, would allow one to tell the differences among sirens (as with violins and cellos). For most applications it should be adequate for the system to provide directional processing from 300 Hz to 3,000 Hz, and omnidirectional detection processing from 10 Hz to 3,000 Hz. It would likely be impractical to build a pole mounted array large enough to provide directional discrimination below 100 Hz (a 3 meter array).

Microphones 500-2, 500-3 and 500-3 are shown connected via electrical delay means having a predetermined delay to summation step 510. In this embodiments, the delay means can be used to permit the circular array to act as a linear array. The delays are adjusted so that the sum of signals for an impinging sound wave is maximum for a signal approaching from a given bearing and less from all other bearings. The delay function in each embodiment may be carried out via programmable hardware and/or software components as are known in the art.

If two poles are provided, each equipped with at least one azimuthal beam-forming array (circular or otherwise) triangulation may be used to obtain x,y coordinates for a sound-producing event occurring in the acoustic coverage area of the two arrays. The area protected by the present system is assumed to be planar, and hence, a reference two dimensional coordinate system (x,y coordinate system) may be used to designate any unique location in the protected area. Of course, the distance between the two poles is known (for example, according to Figure 8, at a distance of four hundred feet). An azimuthal bearing is calculated for each array and the event is assumed to occur at the intersection of the bearings in the x,y coordinate plane of the protected area.

Referring first to Figure 5a, there are shown sound waves approaching from the right in line perpendicular to microphones 500-1 to 500-5. Depending on proper adjustment of the

delays, it may be seen that the expected output of summation step 510 is a relatively large amplitude signal. The gain of the array is set intentionally great or a maximum for signals impinging from the right. The values of the delays are set according to the speed of sound and the distance according to a given bearing (such as 90 degrees) between microphones. Delay 506 is set approximately twice delays 505 and 507 (which are equal) and no delay is needed for phones 500-1 and 500-5. The value of the delay 506 may be calculated as the length of time required for sound to travel from phone 500-3 to either phone 500-1 or 500-5. The value of delays 505, 507 is calculated as the length of time required for sound to travel in the 90 degree bearing direction shown from phones 500-2 and 500-4 to phones 500-1 and 500-5. This process is repeated for a plurality of delay sets to create a plurality of azimuthal beams. It is then determined which beam contains the strongest signal of the candidate event.

Referring now to Figure 5b and assuming the same values for delays 505-507, the sound waves are seen approaching from a 45 degree angle. Consequently, the output summed waveform should have a relatively low amplitude compared with the gain value for Figure 5a. Clearly, a linear array has been approximated from a circular array, the circular array cross-section having a structure that may most conveniently be mounted to a pole (as shown in Figure 7). The circular arrays may be, as already described, stacked eight high to simultaneously form vertical arrays equally spaced about the pole 701 and capable of producing elevational beams.

Figure 6 is a figure demonstrating, in a circular array cross-section, the formation of a horizontal aperture via circular array microphones 600-1 to 600-8. As in Figure 5a, it will be assumed that sound waves are approaching perpendicular to the circular array at microphone 600-3 first. From the arrangement of the delay means (not shown) and the microphones, it can be seen that a pseudo-linear array is approximated by the circular array at the center of the array. The "pseudophones" comprise phones 601, 602 and 603 which do not exist in fact but exist in the equivalent horizontal aperture. Pseudo-phone 603 equates to actual phone 600-4 and its associated delay means, 602 to 600-3 and so on. In this manner a circular array is made to approximate a linear array for receiving and beam forming in a horizontal plane in an equivalent manner to having a plurality of linear arrays. The circular array forms a horizontal aperture and the stacking of such circular arrays, for example, eight high, forms a plurality of vertical linear arrays. To form a two dimensional beam, the data within the horizontal depression angle beam is processed to form azimuthal beams.

It should be noted that in each of the embodiments of Figs. 5-6, the microphones may be directional or omnidirectional. In the former case, the microphones should be directed radially outwardly. In the latter case, microphone orientation is generally immaterial.

Referring now to Figures 7-8, a practical arrangement of an open area security system according to the present invention will be described. Referring first to Figure 7, it may be assumed that an open area to be protected comprises a Peach Tree Station stop of a mass transit system 700. A train 707 is shown emerging from a tunnel on a track 705 pulling in to Peach Tree Station. It may be desirable to provide acoustic arrays and cameras mounted on the roof of station 706, in its interior (not shown), to poles at a platform 709, and to poles 701, 702, 703, and 704 of a parking lot 707. Such an arrangement provides track security, parking security, and station security. More or less security may be provided as appropriate based on cost and other considerations.

Let us assume that an event has occurred in the parking area 707, the reader's attention should now be directed to box 720 showing an automobile 725 leaving a parking area containing a pole 701. Pole 701 is shown equipped with microphone array 1, for example, comprising eight circular arrays of eight phones each per Figure 5a stacked to form eight vertically oriented arrays, represented 721-1 and microphone array 2 or 721-2. Pole 701 is also equipped with camera 723 and pan/zoom control circuits/motors 722. According to the present invention, the processors 100, 110 of the present invention record events and, with or without the assistance of human intervention, the events can be classified emergency or non-emergency or otherwise classified. An event such as a car leaving the parking lot may be routine or considered an emergency depending on the events preceding its departure.

Referring to monitoring command and control center 740, there are provided monitors 740-1 to 740-15. Selected audio may be listened to, events may be monitored, and the operator may define an alert condition for a particular event in memory, override a decision by the depicted control computer or, otherwise, act according to the monitored event.

Referring again to pole 701 where the camera is mounted on the same pole as two acoustic arrays, the acoustic processor 110 determines the bearing to a sound source and so points the adjacent camera 723 down the calculated line of bearing. Zoom, focus and pan control may be allocated to an operator or carried out automatically. In accordance with known video processing techniques coupled with the acoustic processing outputs, the camera 723 can be caused to follow the movement of a suspect with or without human intervention.

At least two arrays may be used to determine by triangulation the x/y coordinates or location of the event, so the camera can be pointed, zoomed and/or focused. The two arrays can be on separate poles in which case standard triangulation can be used to obtain the location coordinates. When the two arrays are on the same pole as shown, then a type of vertical triangulation (cross correlation between the top beam of array 721-1 and the bottom beam of array 721-2 that are on the same bearing) is used to obtain the coordinates.

As already briefly described above, two circular arrays on poles separated by a predetermined distance such as four hundred feet may be used to determine both the bearing to, and x,y coordinates of, events occurring in the acoustic coverage region.

Alternatively, the location (x/y coordinates) of an event can be determined with a single phone array (e.g., cylindrical array) capable of forming both elevational and azimuthal beams. In this case, a depression angle determined from the elevational beams is used in conjunction with a known elevation of the array above the plane of the protected area to compute a range. Azimuth can be determined from the azimuthal beams. Location determination using this technique has limited accuracy, particularly as the distance of the event from the array increases and the depression angle becomes small.

The system of Figure 7 is not shown to suggest that a portable system may not be provided. For example, a camera, a beam forming array and light may be provided on a telescopic pole mounted to a vehicle than can be moved to a location of an event. For example, such a portable system may be used to provide security at a golf club for spectators of a golf tournament or the like, for example, one such vehicle per acre of parking facility.

Most parking lot lights are sodium vapor lights which translates to approximately 200 foot of illumination from a thirty foot pole. For incandescent lights, poles should be mounted more closely together. For example, in a metropolitan area, poles may be spaced at 100 feet along streets and provided with conventional incandescent lighting. Camera systems may be supplemented with infrared illumination or tracking if lighting is insufficient for identification. In the dark, infrared may be preferable for tracking suspect activities. Lasers are getting less expensive over time and laser range finding could provide a means of range finding that is superior to acoustic triangulation processing. Fusion of audio, video, laser and other sensor systems will provide a more complete picture of an event than each sensor system can provide alone.

Human identification is possible at distances of up to 300 feet from a camera with reasonable zoom capability, (two to eight times power). Acoustic detection with a small

array seems to be optimized at this 300 foot maximum distance with 10 degree beam separations.

In Figure 8, a typical open area (a parking lot) security system is shown. Poles 701 and 702 may conveniently comprise 20 foot high poles (although the range for such poles may be from 8 feet to 50 feet). As a trade-off between desired resolution, sound gathering and the like, it is suggested that the poles be placed 200 feet from the perimeter of the parking lot and, for poles mounted in the center of a 400 foot by 800 foot area, approximately 400 feet apart. In this manner, two poles each equipped with only one camera and vertical and horizontal arrays can cover a 400 foot by 800 foot area. Camera systems are commercially available for outputing a wireless 9600 baud (roughly 3000 Hz) audio signal and provides for video data and control for an economical arrangement for communicating with a computer work station 800. Consequently, a very economical arrangement may be provided in accordance with the present invention for monitoring an existing parking lot with little modification. Power should be available at poles 701 and 702 for powering lights and wireless communication eliminates any need for running conduits or designing other expensive area wiring systems.

In another embodiment, each of the two poles may be equipped with circular arrays of eight microphones each (and no vertical arrays) such that by triangulation coordinates for events within the acoustic coverage region of poles 701 and 702 may be easily determined. A shortcoming of this arrangement is that events occurring at x,y coordinates for locations 803 and 804 (or in the vicinity of lines 805, 806) will have a large margin of error without vertical array collected data from the closest pole.

A system according to the present invention may be considerably more complex. According to Figure 9, there are several sector command and control centers represented by sector centers 910-1 to 910-5. Each of the sector centers 910-1 to 910-5 is similarly equipped and may be described similarly as sector center 910-1. Sector center 910-1 comprises a plurality of monitors 901-1 to 901-15 for viewing events. Control computer 902 provides for audio selection, event video monitoring and event definition as already described in connection with center 740 of Figure 7. Each of the sector centers communicates with a central command and control center 920. To do so, each sector center further comprises audio and video compression encoding circuitry, such as MPEG 2 circuitry, digital storage 904 for storing digital audio and video data and network control and modulator circuitry 905 for selecting telecommunications or cable television channels to central center 920.

Central center 920 comprises corresponding receiver/demodulator circuitry 925 at the end of sector communications links indicated as Sector One through Sector Five links by way of example. The received audio and video data which is compressed is decoded and decompressed via circuits 923 and output to control computer 922 and/or stored via digital storage 924. Control computer 922 performs similarly to sector centers 902 for selecting audio, monitoring events and defining alerts. Monitors 920-1, 920-6, 920-11 may be utilized for viewing events related to sector 1 and other monitors arranged by sector for event viewing as suggested through monitor 920-15 for viewing events at sector five.

Figure 10 is a combination flowchart and summary of the open area security system of the present invention. According to step 1, a plurality of sensors, i.e. a multi-sensor, multi-source surveillance of areas or facilities to be secured comprises acoustic sensors, video camera sensors, infrared illumination and sensing, laser illumination and sensing coupled with human observations for evaluating events occurring within the area. Each of these is provided according to a design particular to a given area. Nevertheless, it is a principle of the present invention that at least one camera and one acoustic array be provided in an complex adaptive processing system according to the present invention. For example, the camera activity may control and learn from the acoustic processing activities, and the acoustic activity may be controlled by and learn from the video activity.

In step 2, there is the detection of significant events within the surveillance area. Thereunder, there is listed signal processing activities which may be audio or video but in either event there may be accomplished complex adaptive processing for detection and categorization of the events individually and in combination. In either audio or video processing, there may be motion/non-motion determination. For example, in video processing, stationary objects may be differentiated from moving objects and moving objects identified as to direction and distance and even recognized. Similarly, through audio processing according to acoustic array processing beam forming fundamentals in horizontal and vertical planes, bearings, distance and movement may be determined and objects recognized by their sounds. In particular, there may be video or audio pattern recognition, for example, audio pattern recognition per Figures 11-13. Sound patterns and image patterns may be processed to eliminate noise or other objects respectively which detract from the recognition of the true sound pattern or image.

According to step 3, the various sounds or images of significant events are classified that have been detected by the sensor suite described above in connection with step 1. One

set of classification codes that may be used may comprise binary representations for suspicious, friendly and hostile. A table (or multiple tables) may be formed in memory of events with human intervention associated with particular images, image sequences, sound patterns and sound pattern sequences. The table, as will be described further herein will comprise links from an event or sequence of events to event classification code and to a code related to a recommended course of action. Tables may be compared, for example, and accumulated. A friendly classification of a car leaving a parking lot may be coupled with a hostile classification for a robbery or mugging event to result in an overall hostile classification for the combination of events. The result of the combination may be the video tracking of the car leaving the parking lot (if hostile) when, otherwise, the system would not track the car (if friendly).

Generally, in a complex adaptive processing system according to the present invention, and according to step 5, there must exist information management to develop and maintain a composite picture (both acoustic and video) of the surveillance area. As described earlier, one event may be associated with another event in memory (a woman's screams with a car leaving the parking lot) so as to result in an overall evaluation of an overall event. The sensors themselves must be manually or automatically managed to provide appropriate inputs in line with events. For example, the camera must be zoomed to capture the image of an attacker. The higher resolution image resulting therefrom can be post-capture processed to determine the attacker's approximate weight, height, and other descriptive information. Correlation is the concept of correlating stored event data with actual event data and their classification. Resource allocation relates to the concept of proper and economical allocation of computer and sensor resources to design an appropriate security system for a given area according to the herein described principles. Fusion is the concept that each of the acoustic and video imaging systems is not a stand-alone system but one aids the other and taken together provide a far better picture of the surveillance area than either alone. Decision aids is the concept that the present invention in providing a complete picture provides a decision aid to a security person manning a command and control center. The collected data may point to certain automated acts done without human intervention, such as pointing a camera toward movement or toward a sound classified as hostile. Yet, there may be provided additional alternative choices for human acceptance or rejection such as recommendations to contact an emergency system for dispatching an emergency vehicle. A screen of the control

center, for example, may be a touch screen whereby automatic dialing of a rescue squad telephone number may be actuated.

Finally in step 6 there is shown the process of action recommendation and selection to maintain security in the surveillance area. With human automated or intervention, events and sequences of events can be linked (for example via linked memory addresses) to actions and recommendations for actions which may be selected but otherwise not automatically engaged. The kind of actions include monitoring events, engaging an attacker, doing nothing, investigating by means of image and sound pattern processing, attacking through alarm systems or the dispatch of personnel to the scene, reporting to a central command or to responsible emergency agencies and diverting or challenging the event (is it truly an emergency as suggested by the system?).

Figures 11-13 are typical sound patterns collected for typical and atypical events that may be stored in memory, individually classified and, moreover, classified if they occur in sequence and tabulated. Figures 11 to 13 show amplitude in a vertical dimension versus frequency from 0 to 500 Hz over time in seconds. The illustrated 0-500 Hz frequency range is merely representative. A spectrum larger than 500 Hz may be accumulated since sound signals may be collected at frequencies up to and in excess of 20,000 Hz.

Figures 11-13 all assume a frequency range of interest at from 0-500 Hz. Figure 11 shows a spectrum for a pistol firing. A pistol firing exhibits a spectrum which, over a two second span, reaches triangular peaks and recedes (between two and four seconds). There are shown relatively high peaks at certain low frequencies such as about 90 Hz, 200 Hz, 300 Hz and 400 Hz. A rifle will exhibit a different but similar spectrum. In fact, different pistols having differently dimensioned ammunition will exhibit different spectra. The collection of such data may lead to the identification of a particular weapon used in an assault.

Figure 12 shows a spectrum for a woman screaming. Notice that there is a strong peak at the frequency of the scream, approximately 400 Hz (higher frequencies are not shown). There are no substantial peaks below 400 Hz. A woman screaming prior to a pistol firing may represent a sequence of events that is clearly classified as "hostile". The occurrence of a woman screaming after a pistol firing may likewise be classified as hostile but likely lead to the inference in a complex adaptive processing system according to the present invention that the screaming woman is not a victim of a bullet wound.

Figure 13 shows a 0-500 Hz spectrum over 0-50 seconds time for a train inbound and outbound. The outbound train includes a greater volume of sound at approximately the same frequencies between 100 and 200 Hz (because the train's engines are running compared with coasting in or braking sounds).

It is known in the telecommunications arts that sounds in the range 0-3000 Hz are useful in providing intelligibility of speech. Sounds above 3000 Hz are not generally carried and, in fact, a so-called C-message (attenuation) curve is applied to sound signals such that signals at 1000 Hz are not attenuated to preserve intelligibility. Moreover, known wireless and wired communications systems typically provide 9600 baud or analog 3000 Hz communications channels. Consequently, it is recommended that to, for example, capture and store an entire conversation, for example, between an attacker and a victim, it is useful to provide arrays which are designed for a 0-3000 Hz spectrum and not simply the 0-500 Hz spectrum shown in Figures 11-13.

Now, a simple scenario will be described in connection with a discussion of Figure 14a. An attack begins, a woman screams. Automatically, an acoustic array forwards a summed horizontal and vertical array signal for acoustic processing. If an array capable of beamforming in either plane captures the sound, not just a direction but x,y coordinates (in the plane of the protected open area) of the attack may be determined. If one array capable of only azimuthal beamforming detects the attack, a camera mounted on the pole with the array may be automatically pointed, but not precisely zoomed or focused. An operator may be present to assist with focus and zoom. Yet, a camera mounted on a twenty foot pole will not lose much detail, even if the attack is immediately below the pole (or at a distance of twenty feet). With video signal processing (via movement detection), the camera can be automatically be focused and zoomed. Or, in the alternative, with two arrays providing data and the x,y coordinates calculated, the camera can be automatically focused and zoomed.

Referring to Figure 14b, there may be prestored in memory a table including event data for a particular event (such as a woman's scream), a binary code representing a preliminary classification of the event (for example, 01 stands for emergency, hostile) and a pointer to another table showing automatic and recommended courses of action. For example, the pointer 0110 may comprise a pointer to a table including automatic actions of pointing a camera in the direction of the bearing and recommended actions which may be displayed to the operator such as recommending that assistance be dispatched. The binary

values and events are merely suggestive of codes and pointers that may be provided. Similar tables may be constructed for video processing activity.

Thus there has been shown and described a system and method for monitoring and securing an open area. Other embodiments and modifications of the described embodiments may have already come to mind. The patents and publications mentioned herein should be deemed to be incorporated by reference herein as to any subject matter believed to be essential to an understanding of the present invention. The invention should only be considered to be limited in scope by the claims that follow.

WHAT WE CLAIM IS

1. A system for monitoring an area, the system comprising:

at least one acoustic sensor array of microphones,

a sound signal processor, coupled to the at least one acoustic array of microphones, for processing received sound inputs and localizing the direction from which the sound input originates and

at least one camera for capturing images, the sound processor for controlling the camera to be directed according to the determined direction.

2. A monitoring system as recited in claim 1 further comprising a memory, coupled to the sound processor, for recording input sound patterns including frequency response data recorded over time for the recorded sounds.

3. A monitoring system as recited in claim 1 wherein said microphones are coupled to said sound processor via a wireless communication link.

4. A monitoring system as recited in claim 2 wherein said memory of said processor further stores classification data for input sound patterns and outputs the classification data regarding a received sound input as to its emergency or a different classification.

5. A monitoring system as recited in claim 4 wherein said memory of said sound processor further stores data regarding one of an automatic or a recommended course of action for input sound patterns.

6. A monitoring system as recited in claim 5, said at least one camera for capturing images wherein said automatic course of action comprises actuating said camera to be directed according to the output direction of the received sound.

7. A monitoring system as recited in claim 2 further comprising a video signal processor for detecting movement and determining direction of said detected movement, said video signal processor, responsive to said sound processor, for actuating said camera to follow the direction of movement over time and to focus according to determined distance.

8. A monitoring system as recited in claim 1 further comprising a second array separated from the first array by a predetermined distance and wherein the output of each array is processed, the processing resulting in an x,y coordinate for an event.

9. A monitoring system as recited in claim 1 further wherein said array forms azimuthal and elevational beams and is situated in known elevational relationship to said area.

10. A monitoring system as recited in claim 1 wherein said camera is mounted on a pole with said acoustic sensor array and is directed according to a determined bearing.

11. A monitoring system as recited in claim 9 wherein said beams are less than 25 degrees wide by less than fifty degrees high.

12. A monitoring system according to claim 1 wherein said array comprises at least a circular array of phones.

13. A monitoring system according to claim 12 wherein said array comprises a plurality of stacked circular arrays.

14. A monitoring system according to claim 13 wherein said plurality of stacked circular arrays forms a cylinder.

15. A monitoring system according to claim 13 wherein said plurality of stacked circular arrays forms a sphere.

16. A monitoring system according to claim 12 wherein each said circular array has a diameter of less than 3 meters.

17. A system according to claim 1 wherein said camera and said array are mounted in an elevated position in relation to said area.

18. A system for monitoring an area, the system comprising at least two acoustic sensor arrays of microphones, each array comprising a plurality of microphones for beamforming, each array being separated by a predetermined distance from one another, each array coupled to a processor for processing received sound inputs and localizing the direction and determining x,y coordinates from which the sound input originates when the sound is input from the area within the acoustic range of the arrays,

said acoustic signal processor and

a camera, responsive to said acoustic signal processor, in known locational relationship to said arrays, said camera for capturing an image at the determined x,y coordinates.

19. A method for monitoring an area comprising the steps of:

storing a plurality of predetermined sound patterns in memory obtained via an acoustic sensor array for forming an azimuthal beam and an elevational beam and associated classification data for each said stored sound pattern,

storing location data regarding said sensor array in relation to said area,

storing location data regarding a camera in relation to said area,

receiving a sound pattern via said acoustic sensor array,

localizing and determining a direction to said sound pattern,

classifying said received sound pattern when said received sound pattern approximately matches one of said stored sound patterns and

directing said camera in said determined direction of said sound pattern.

20. A method according to claim 19 further comprising the steps of simultaneously receiving the sound pattern via a second acoustic array spaced at a predetermined distance from said acoustic sensor array and determining an x,y coordinate for an event initiating said received sound pattern.

21. A method according to claim 19 further comprising the step of storing a course of action for said classified sound pattern.

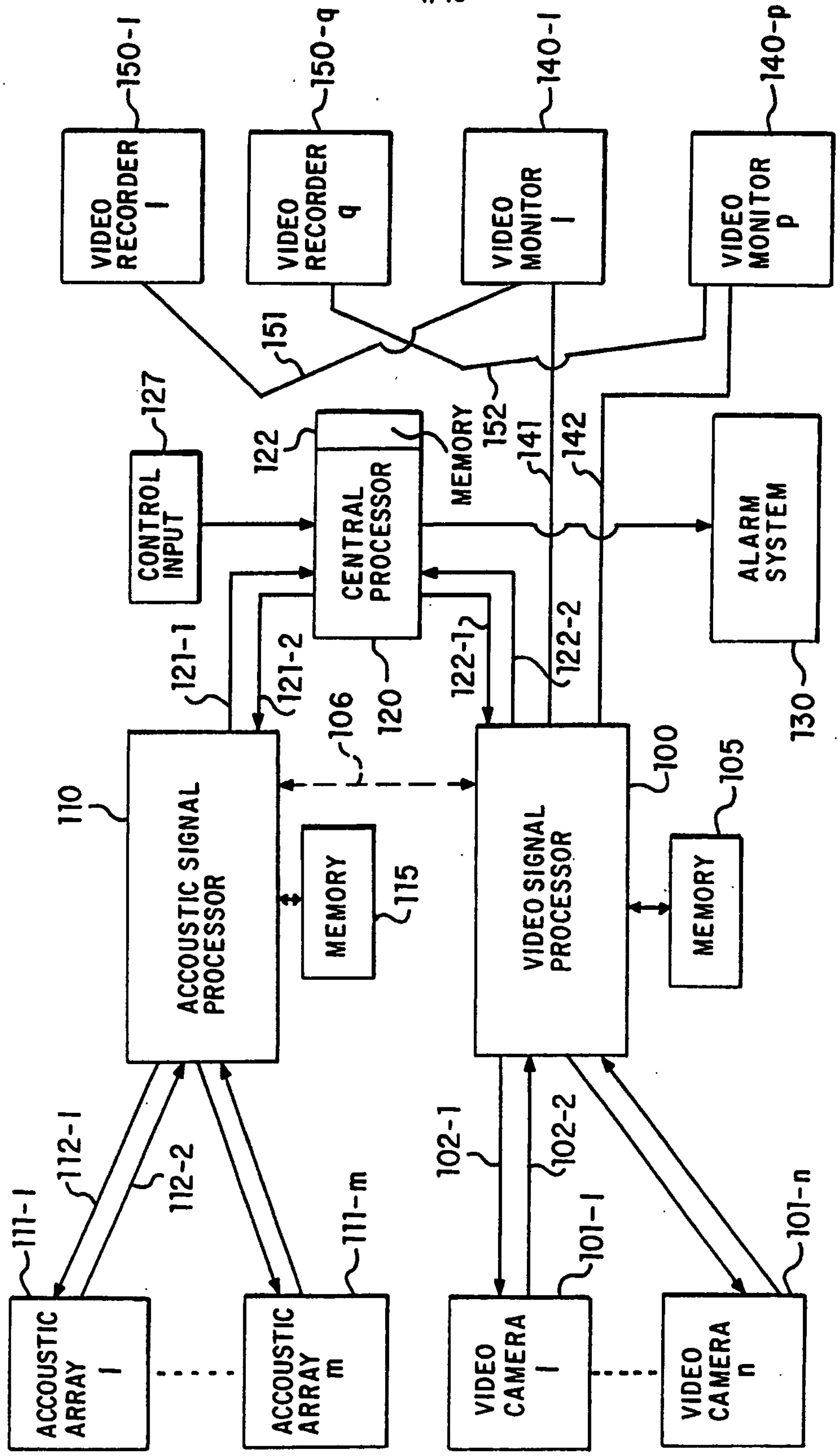
22. A method according to claim 21 further comprising the step of automatically initiating said stored course of action.

23. A method according to claim 22 wherein said direction determining step further comprises calculating x,y coordinates for said classified sound pattern using a predetermined elevational relationship between the location of said array and said area.

24. A method according to claim 20 further comprising the step of directing a camera in the determined direction and focusing said camera at a determined distance from the x,y coordinates.

25. A method according to claim 19 further comprising the steps of storing video data representing an event, determining movement of a portion of the image and zooming and focusing a camera responsive to the movement determination.

26. A method according to claim 19 providing directional processing from approximately 300 to 3000 Hz and omnidirectional classification processing from approximately 10 Hz to 3,000 Hz.



VIDEO RECORDER 1
150-1
VIDEO RECORDER q
150-q
VIDEO MONITOR 1
140-1
VIDEO MONITOR p
140-p
151
152
127
CONTROL INPUT
122
MEMORY
CENTRAL PROCESSOR
141
142
ALARM SYSTEM
130
121-1
121-2
106
120
122-1
122-2
110
ACCOUSTIC SIGNAL PROCESSOR
MEMORY
115
VIDEO SIGNAL PROCESSOR
100
MEMORY
105
112-1
112-2
102-1
102-2
111-1
111-m
101-1
101-n
ACCOUSTIC ARRAY 1
ACCOUSTIC ARRAY m
VIDEO CAMERA 1
VIDEO CAMERA n

FIG. 1

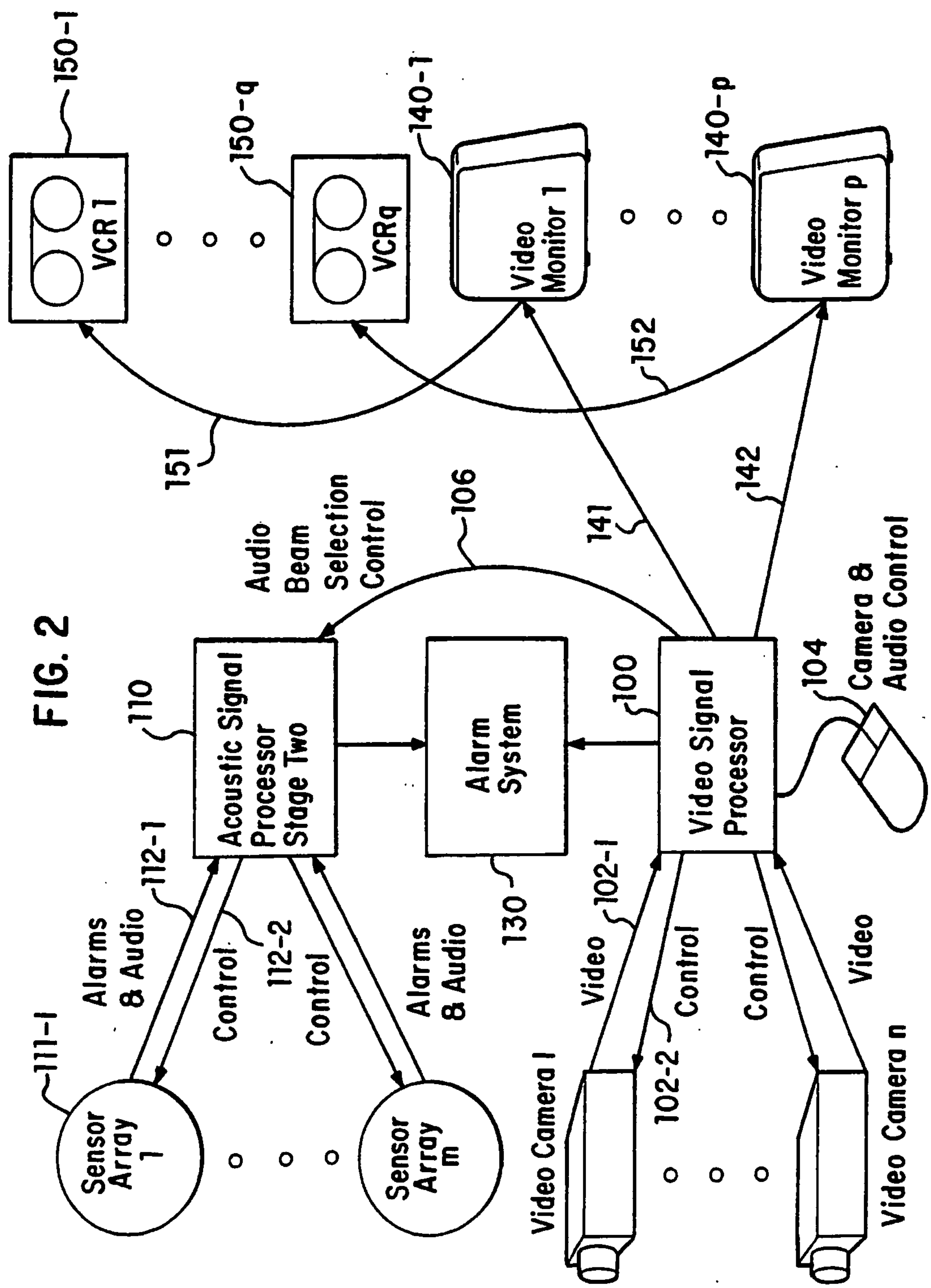
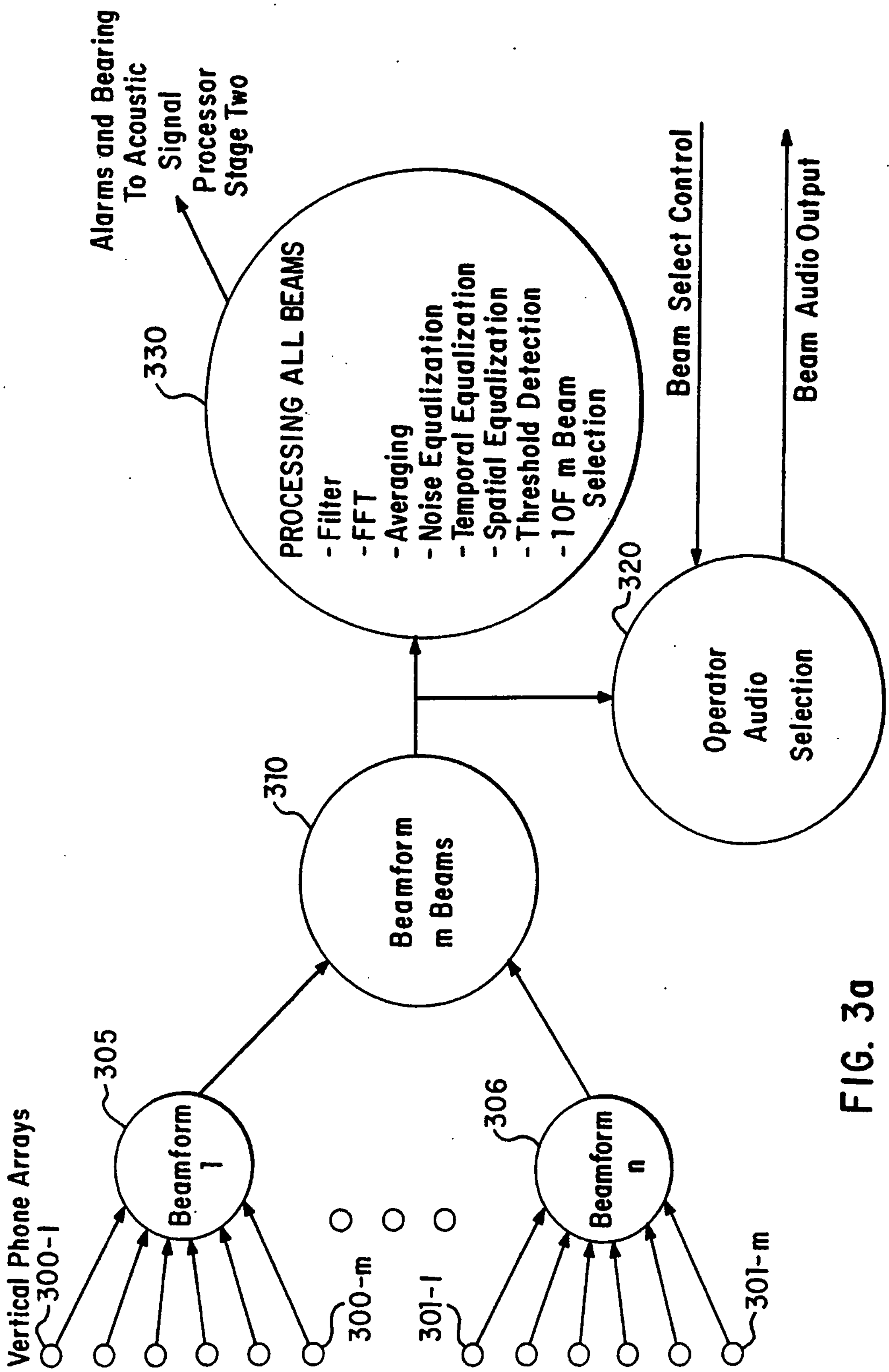

150-1
VCR 1
150-q
VCRq
140-1
Video Monitor 1
140-p
Video Monitor p
151
152
141
142
Audio Beam Selection Control
106
110
Acoustic Signal Processor Stage Two
Alarm System
100
Video Signal Processor
104
Camera & Audio Control
130
111-1
Sensor Array 1
Sensor Array m
Alarms & Audio
112-1
Control
112-2
Control
Alarms & Audio
Video
102-1
Control
102-2
Control
Video
Video Camera 1
Video Camera n

FIG. 2



Vertical Phone Arrays
300-1
300-m
301-1
301-m
Beamform 1
305
Beamform n
306
Beamform m Beams
310
PROCESSING ALL BEAMS
- Filter
- FFT
- Averaging
- Noise Equalization
- Temporal Equalization
- Spatial Equalization
- Threshold Detection
- 1OF m Beam Selection
330
Alarms and Bearing To Acoustic Signal Processor Stage Two
Operator Audio Selection
320
Beam Select Control
Beam Audio Output

FIG. 3a

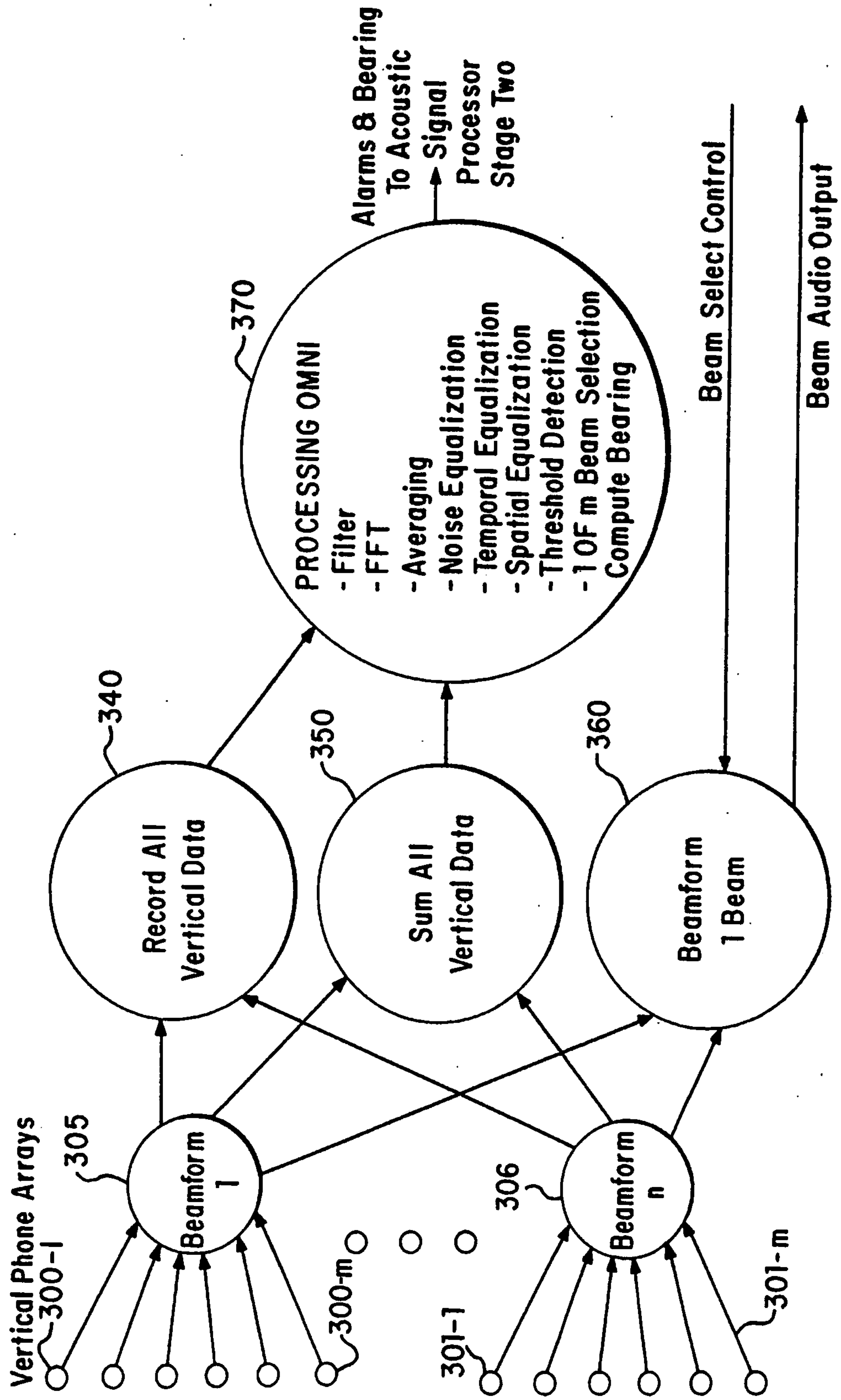

Vertical Phone Arrays
300-1
300-m
301-1
301-m
305
Beamform 1
306
Beamform n
340
Record All Vertical Data
350
Sum All Vertical Data
360
Beamform 1 Beam
370
PROCESSING OMNI
- Filter
- FFT
- Averaging
- Noise Equalization
- Temporal Equalization
- Spatial Equalization
- Threshold Detection
- 1 OF m Beam Selection
Compute Bearing
Alarms & Bearing To Acoustic Signal Processor Stage Two
Beam Select Control
Beam Audio Output

FIG. 3b

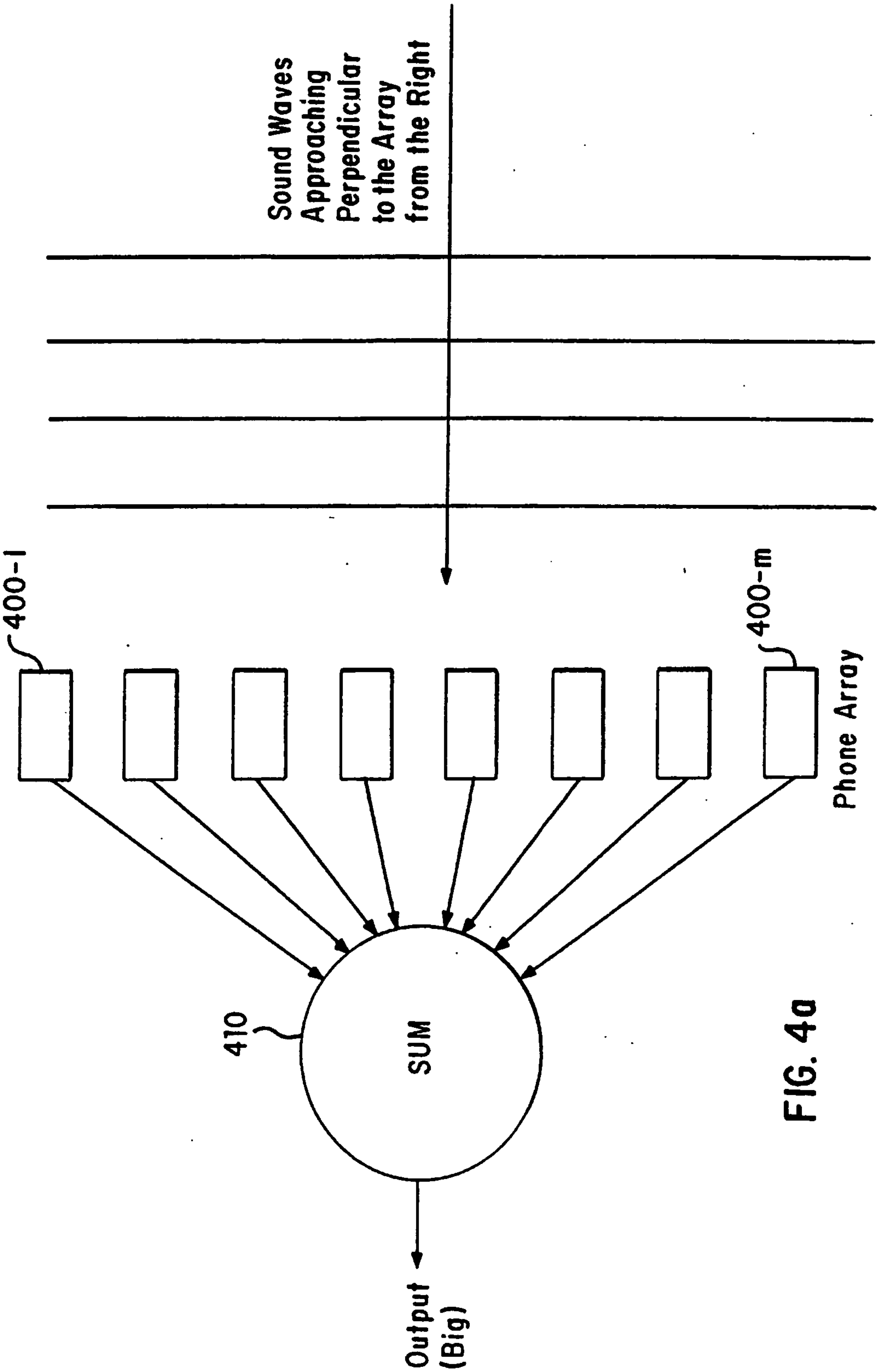

Sound Waves
Approaching
Perpendicular
to the Array
from the Right
400-1
400-m
Phone Array
410
SUM
Output
(Big)

FIG. 4a

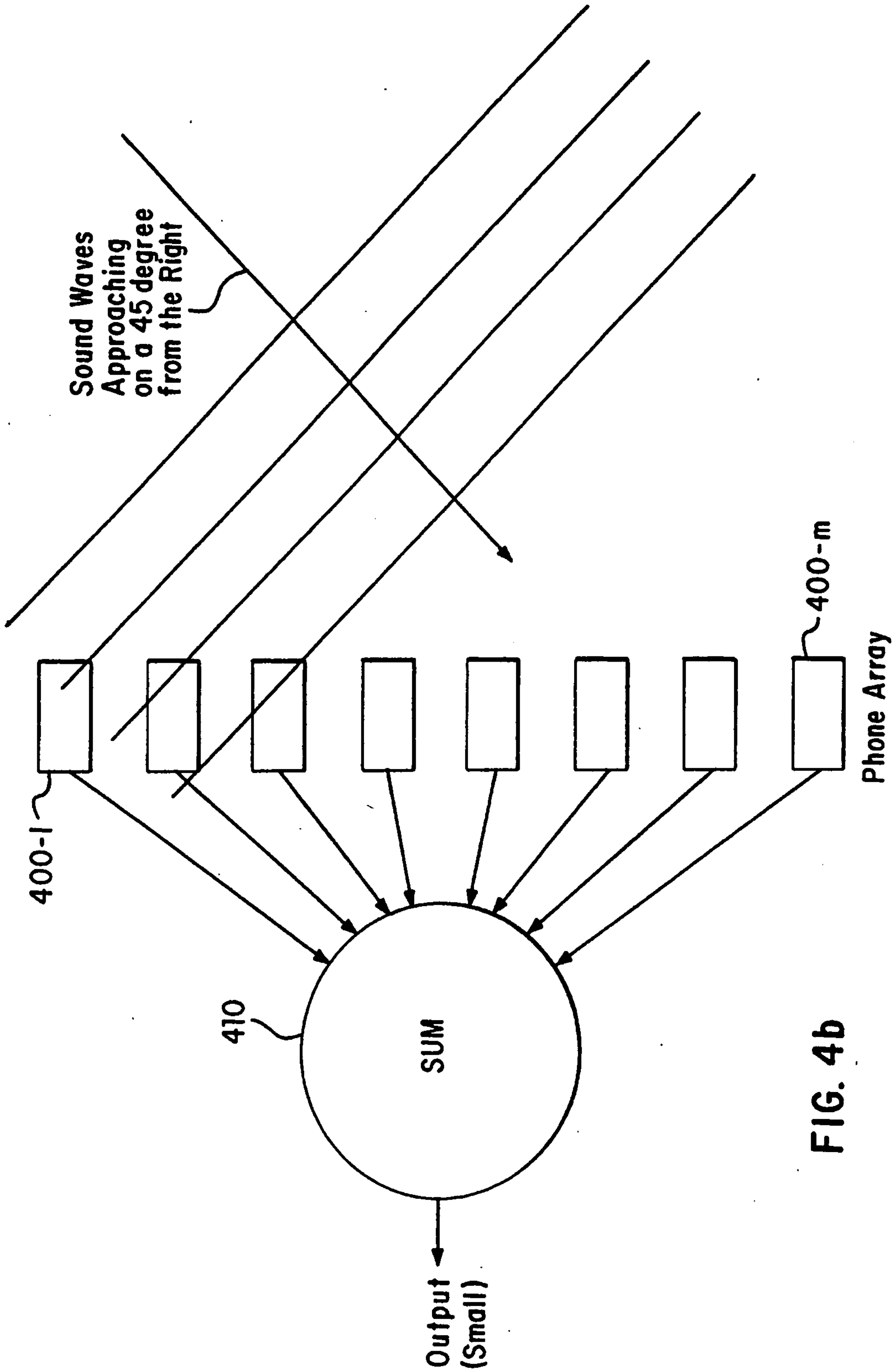

Sound Waves
Approaching
on a 45 degree
from the Right
400-m
Phone Array
400-l
410
SUM
Output
(Small)

FIG. 4b

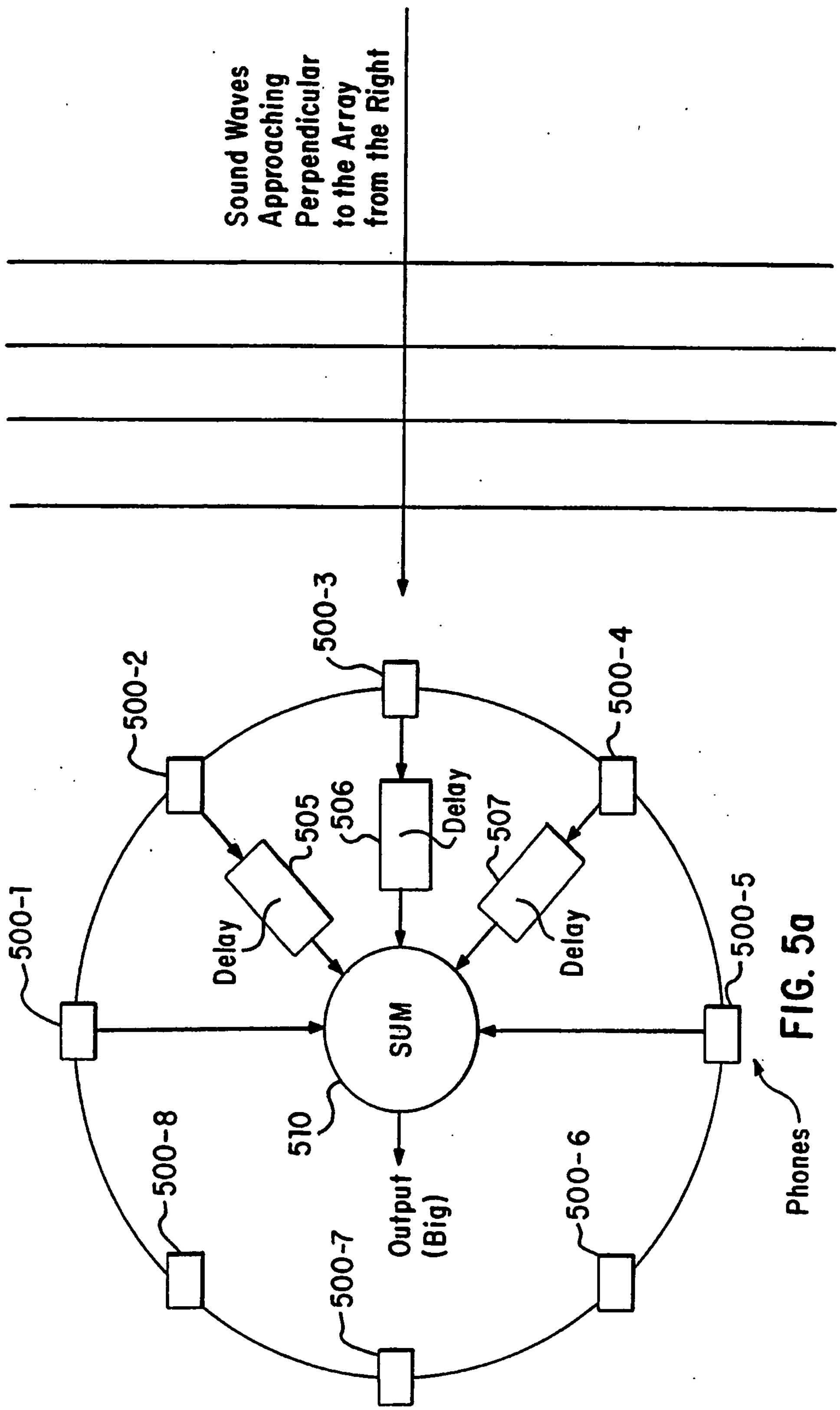

Sound Waves
Approaching
Perpendicular
to the Array
from the Right
500-1
500-2
500-3
500-4
500-5
500-6
500-7
500-8
505
506
507
Delay
Delay
Delay
SUM
510
Output
(Big)
Phones

FIG. 5a

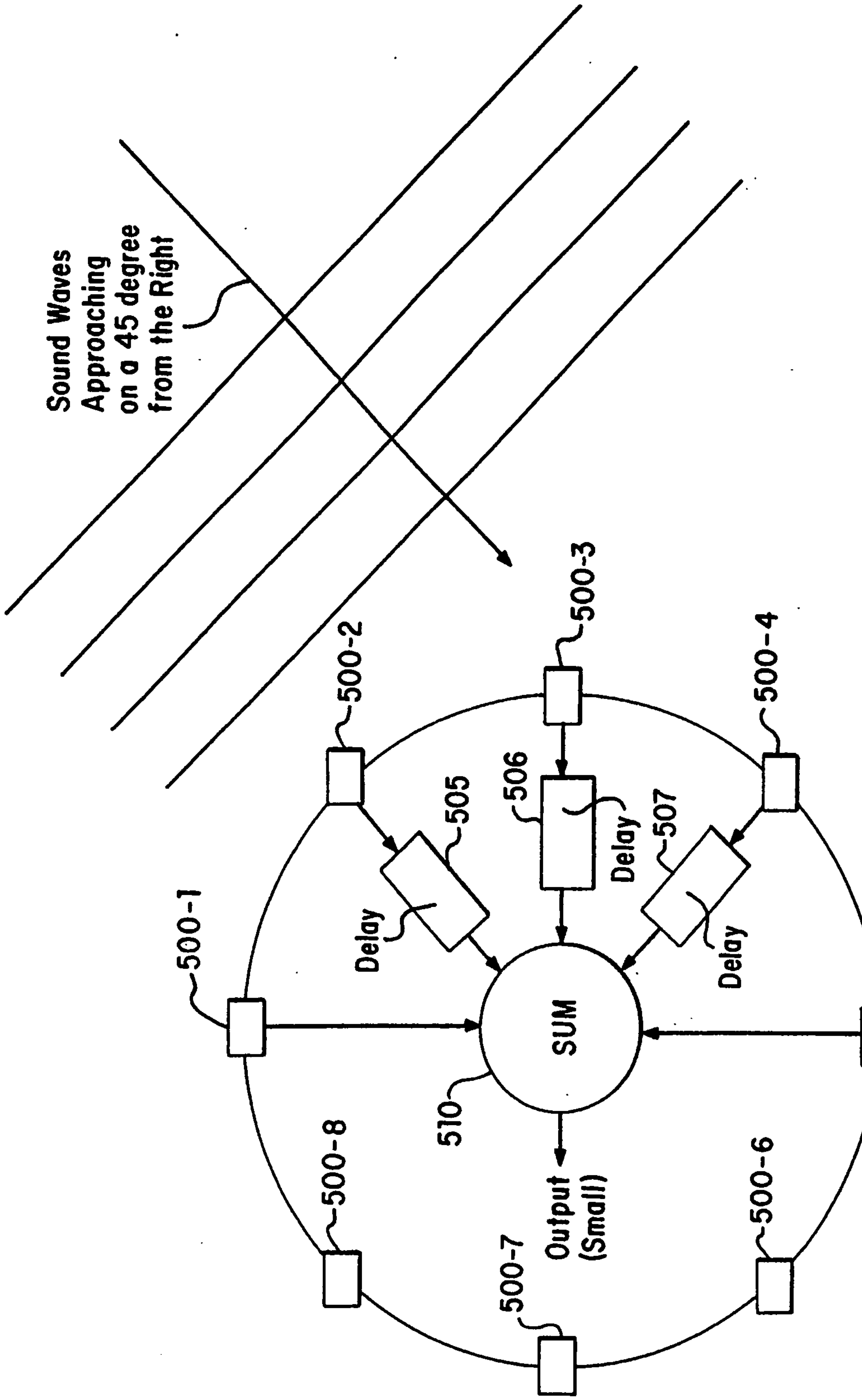

Sound Waves
Approaching
on a 45 degree
from the Right
500-1
500-2
500-3
500-4
500-5
500-6
500-7
500-8
505
506
507
Delay
Delay
Delay
SUM
510
Output
(Small)
Phones

FIG. 5b

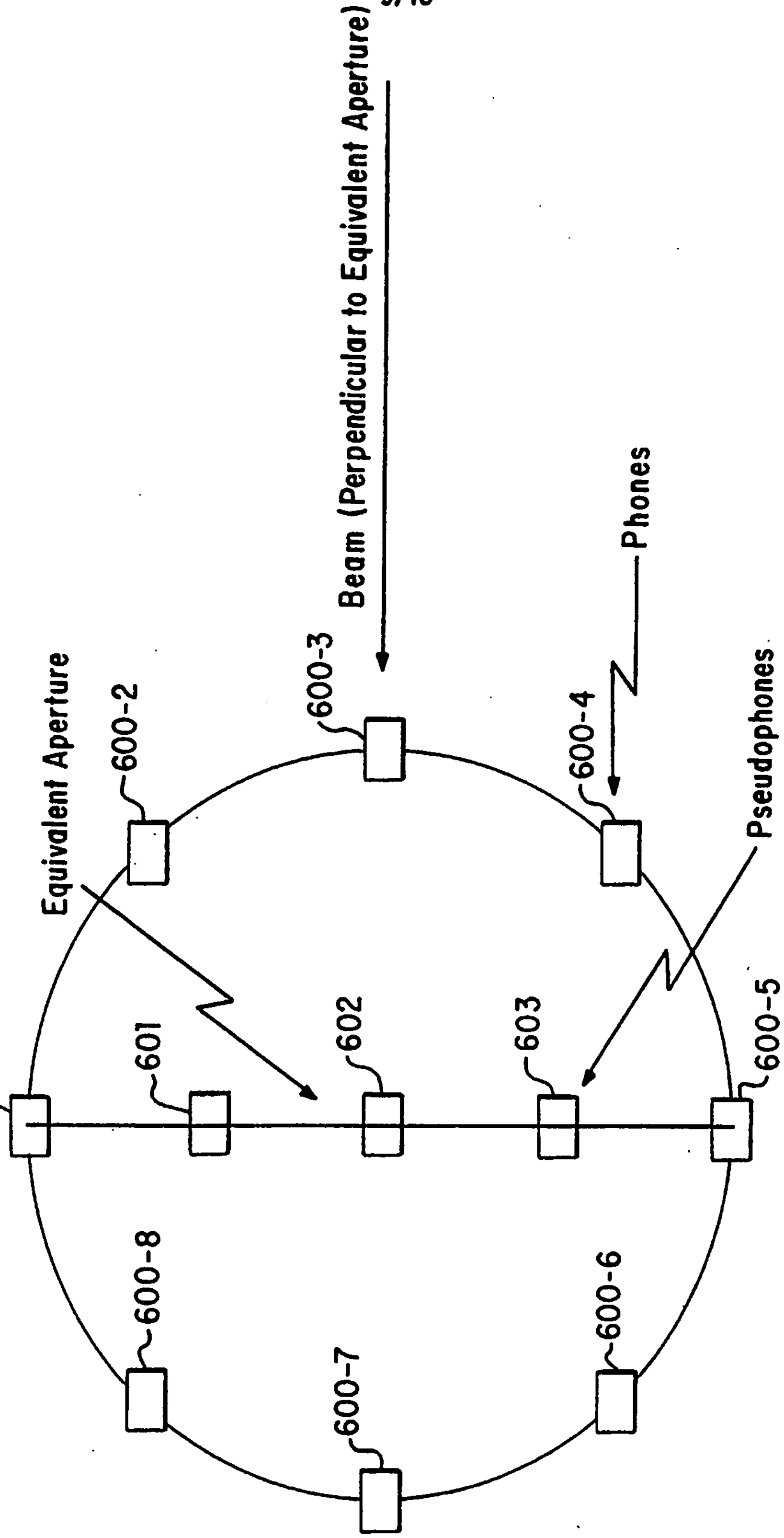

Beam (Perpendicular to Equivalent Aperture)
Phones
Pseudophones
Equivalent Aperture
600-1
600-2
600-3
600-4
600-5
600-6
600-7
600-8
601
602
603

FIG. 6

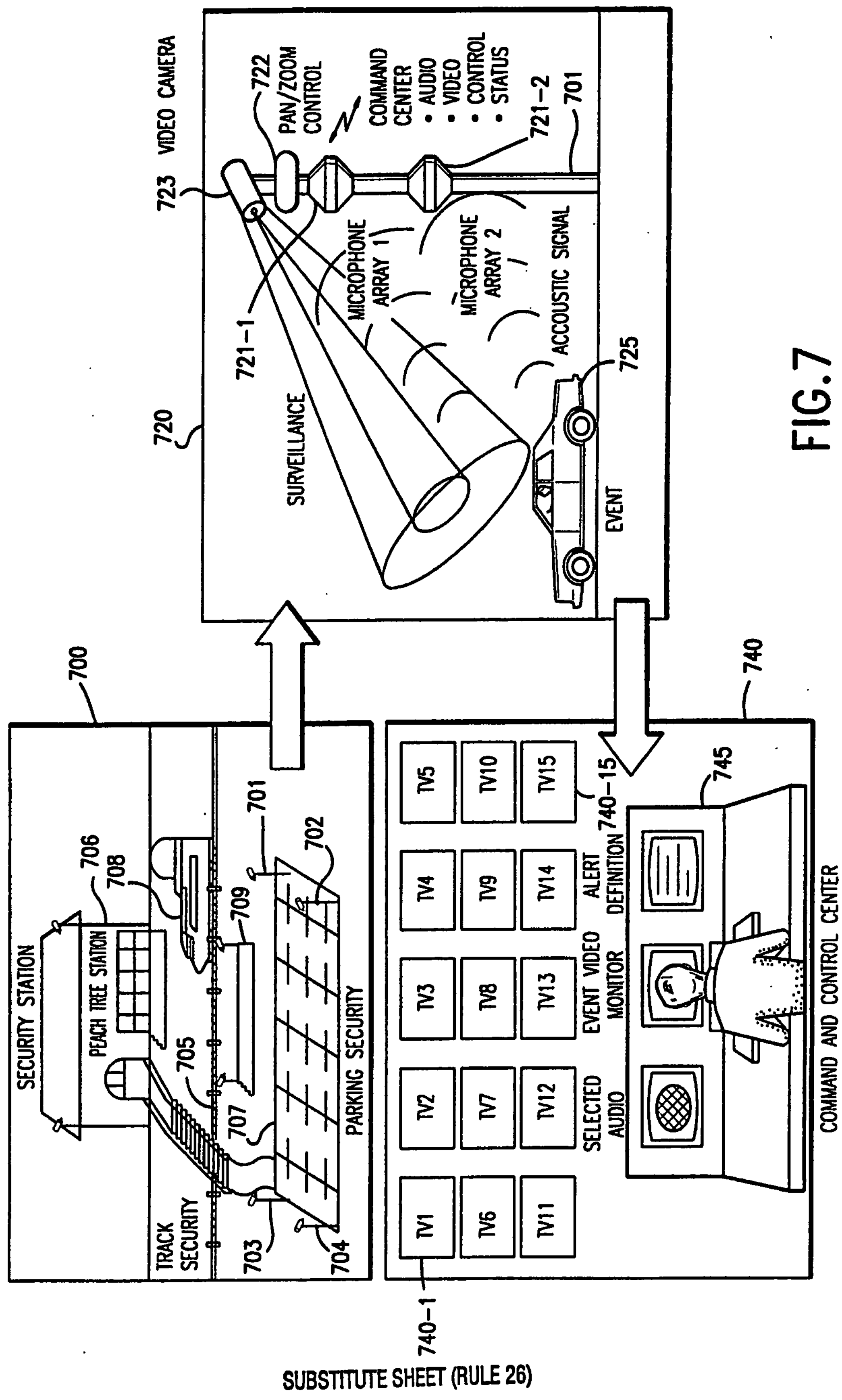

700
SECURITY STATION
PEACH TREE STATION
TRACK SECURITY
PARKING SECURITY
701
702
703
704
705
706
707
708
709
720
SURVEILLANCE
723 VIDEO CAMERA
722 PAN/ZOOM CONTROL
COMMAND CENTER
• AUDIO
• VIDEO
• CONTROL
• STATUS
721-1
721-2
701
MICROPHONE ARRAY 1
MICROPHONE ARRAY 2
ACCOUSTIC SIGNAL
725
EVENT
740
TV1
TV2
TV3
TV4
TV5
TV6
TV7
TV8
TV9
TV10
TV11
TV12
TV13
TV14
TV15
740-1
740-15
SELECTED AUDIO
EVENT VIDEO MONITOR
ALERT DEFINITION
745
COMMAND AND CONTROL CENTER

FIG. 7

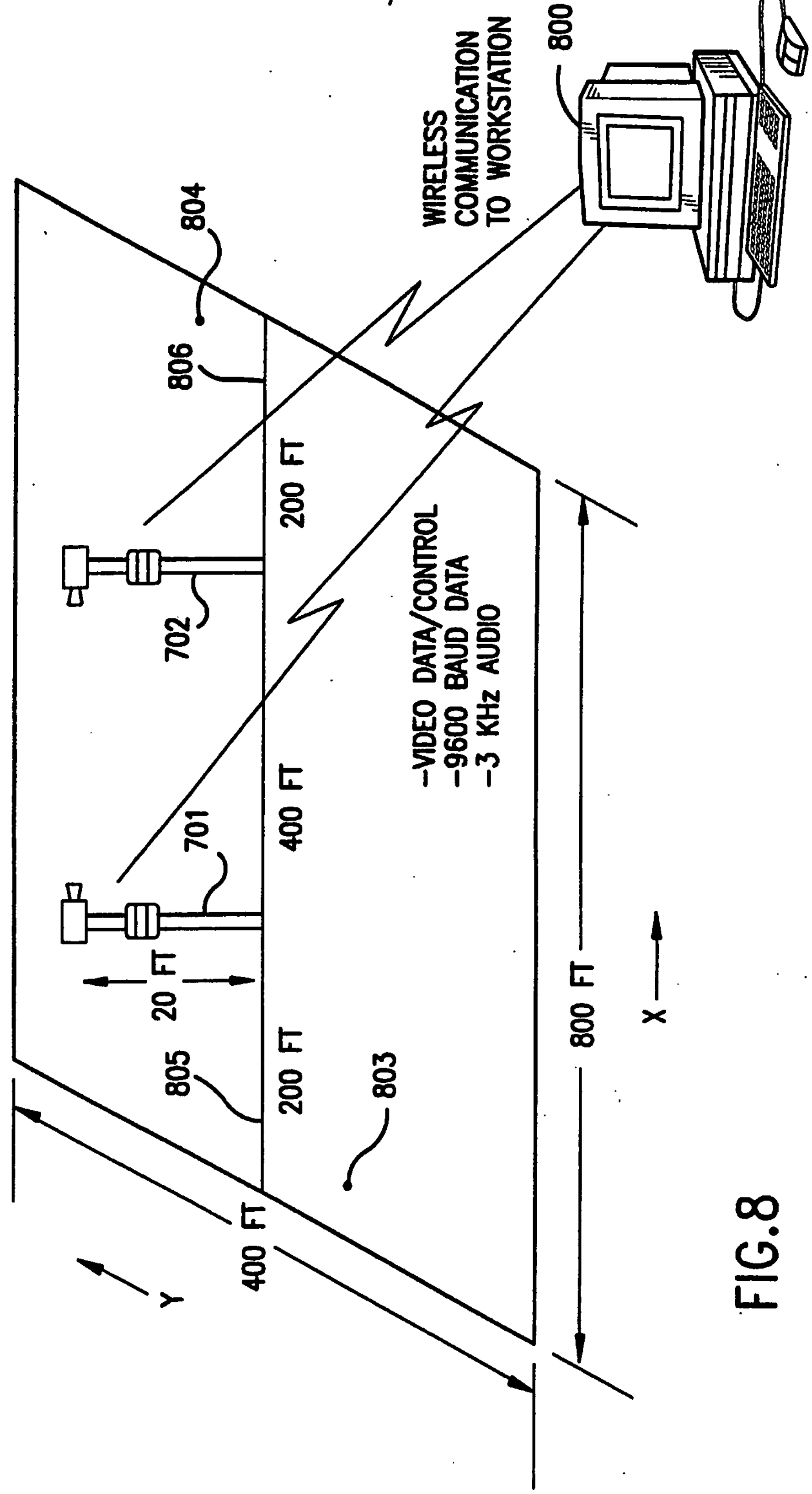

800
WIRELESS
COMMUNICATION
TO WORKSTATION
804
806
200 FT
702
–VIDEO DATA/CONTROL
–9600 BAUD DATA
–3 KHz AUDIO
701
400 FT
20 FT
200 FT
805
803
800 FT
X
Y
400 FT

FIG.8

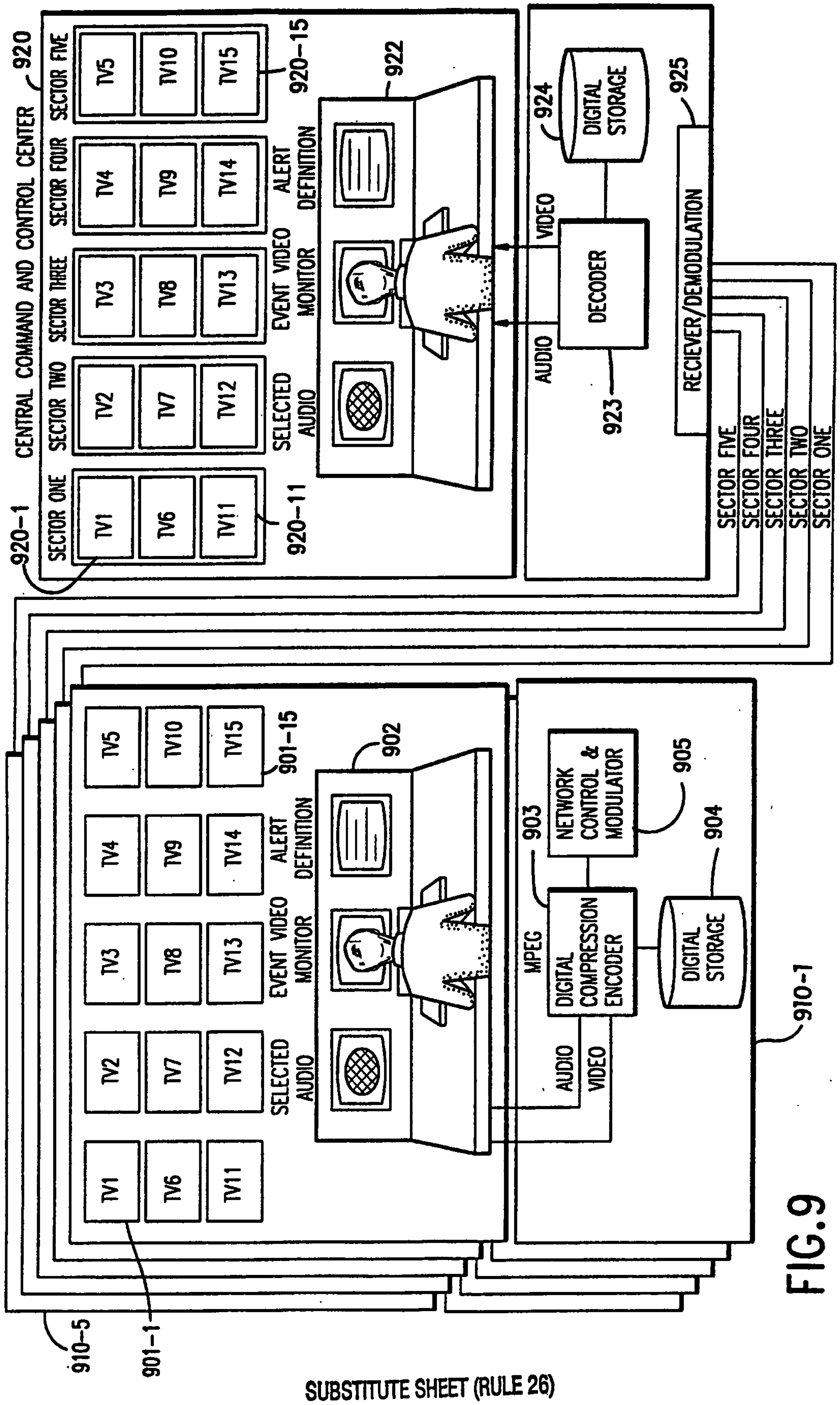

CENTRAL COMMAND AND CONTROL CENTER
920
SECTOR ONE
SECTOR TWO
SECTOR THREE
SECTOR FOUR
SECTOR FIVE
TV1
TV2
TV3
TV4
TV5
TV6
TV7
TV8
TV9
TV10
TV11
TV12
TV13
TV14
TV15
920-1
920-11
920-15
SELECTED AUDIO
EVENT VIDEO MONITOR
ALERT DEFINITION
922
AUDIO
VIDEO
DECODER
923
DIGITAL STORAGE
924
925
RECIEVER/DEMODULATION
SECTOR FIVE
SECTOR FOUR
SECTOR THREE
SECTOR TWO
SECTOR ONE
901-15
902
903
MPEG
DIGITAL COMPRESSION ENCODER
NETWORK CONTROL & MODULATOR
905
904
910-1
910-5
901-1

FIG.9

1. Multi-sensor, multi-source surveillance of specific areas/facilities:

* Acoustic	* Video	* Infrared
* Laser	* Human Observations	

2. Detection of significant events within the surveillance area:

* Signal Processing	* Pattern Recognition
* Motion/Non-Motion Determination	* Image Processing

3. Classification of significant events detected by the sensor suite:

* Suspicious	* Friendly
* Hostile	

4. Localization and tracking of significant events detected and classified as significant:

* Geopositional	* Sector-to-Sector
* Sensor-to-Sensor	

5. Information management to develop and maintain a composite picture of the surveillance area:

* Association	* Correlation	* Fusion
* Sensor Management	* Resource Allocation	* Decision Aids

6. Action recommendation and selection to maintain security in the surveillance area:

* Monitor	* Investigate	*Report
* Engage	* Attack	* Divert/Challenge Event
* Do Nothing		

FIG. 10

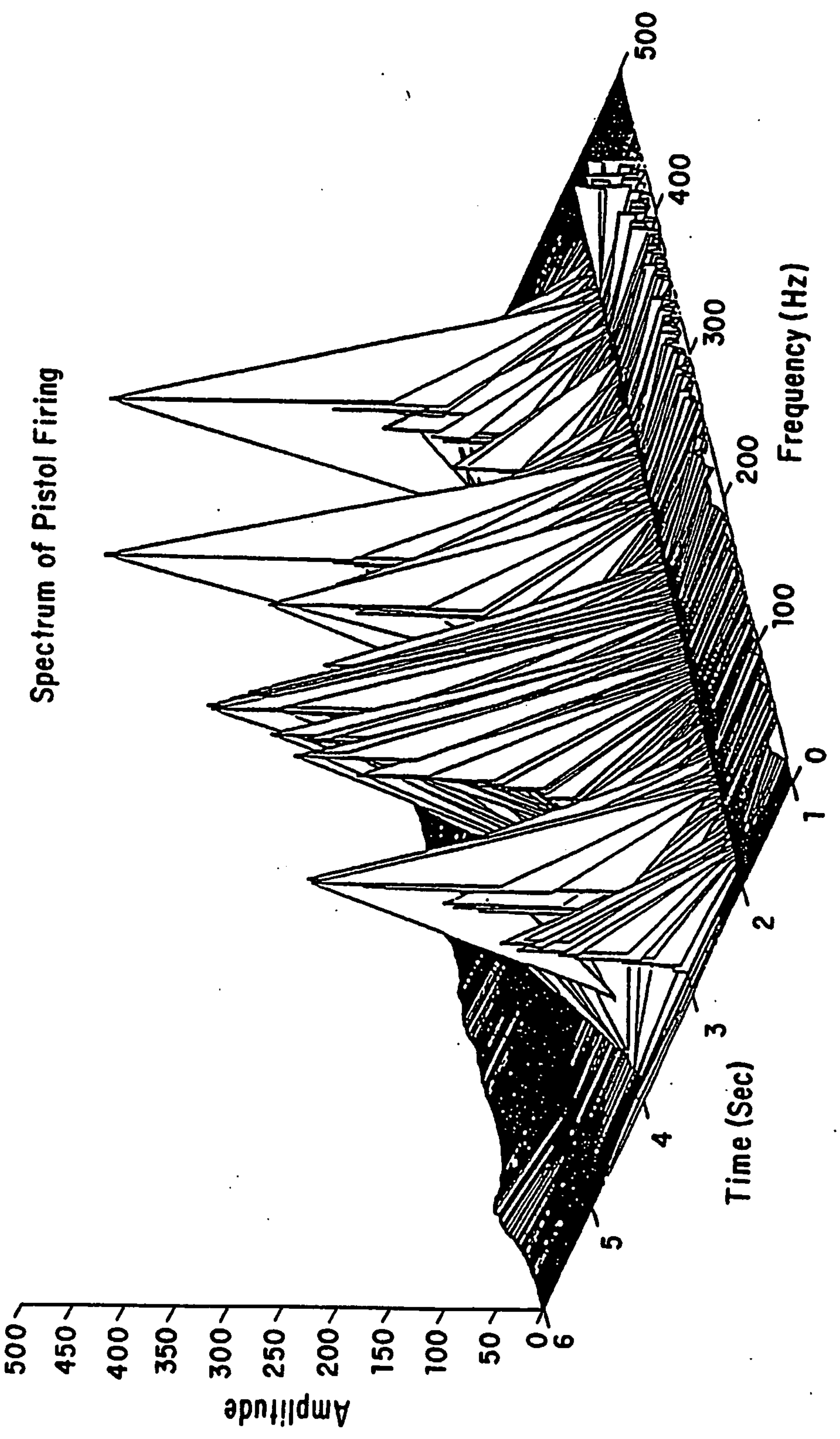

Spectrum of Pistol Firing
Frequency (Hz)
0
100
200
300
400
500
Time (Sec)
1
2
3
4
5
6
Amplitude
0
50
100
150
200
250
300
350
400
450
500

FIG. 11

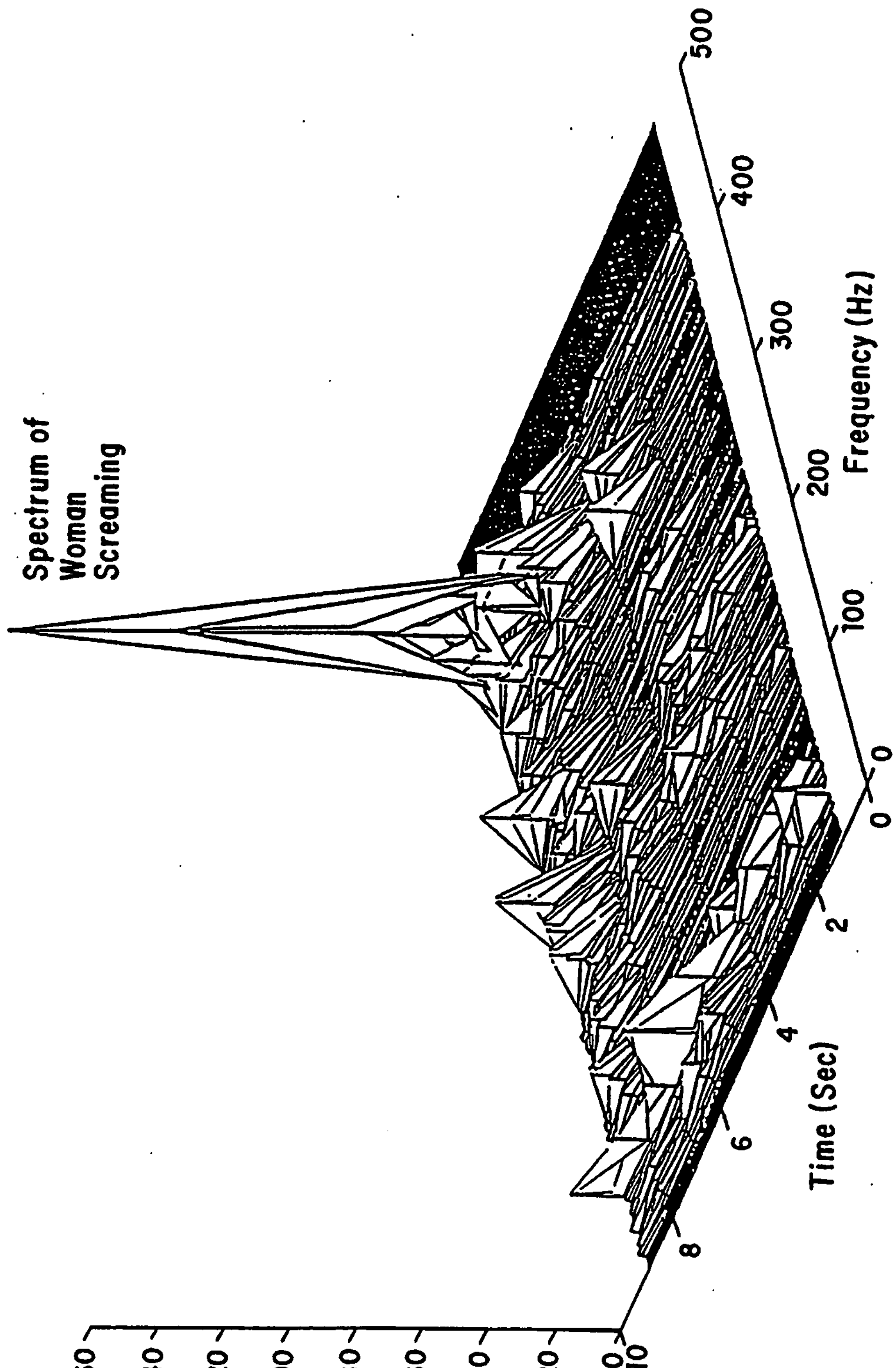

Spectrum of
Woman
Screaming
Frequency (Hz)
0
100
200
300
400
500
Time (Sec)
0
2
4
6
8
10
Amplitude
0
20
40
60
80
100
120
140
160

FIG. 12

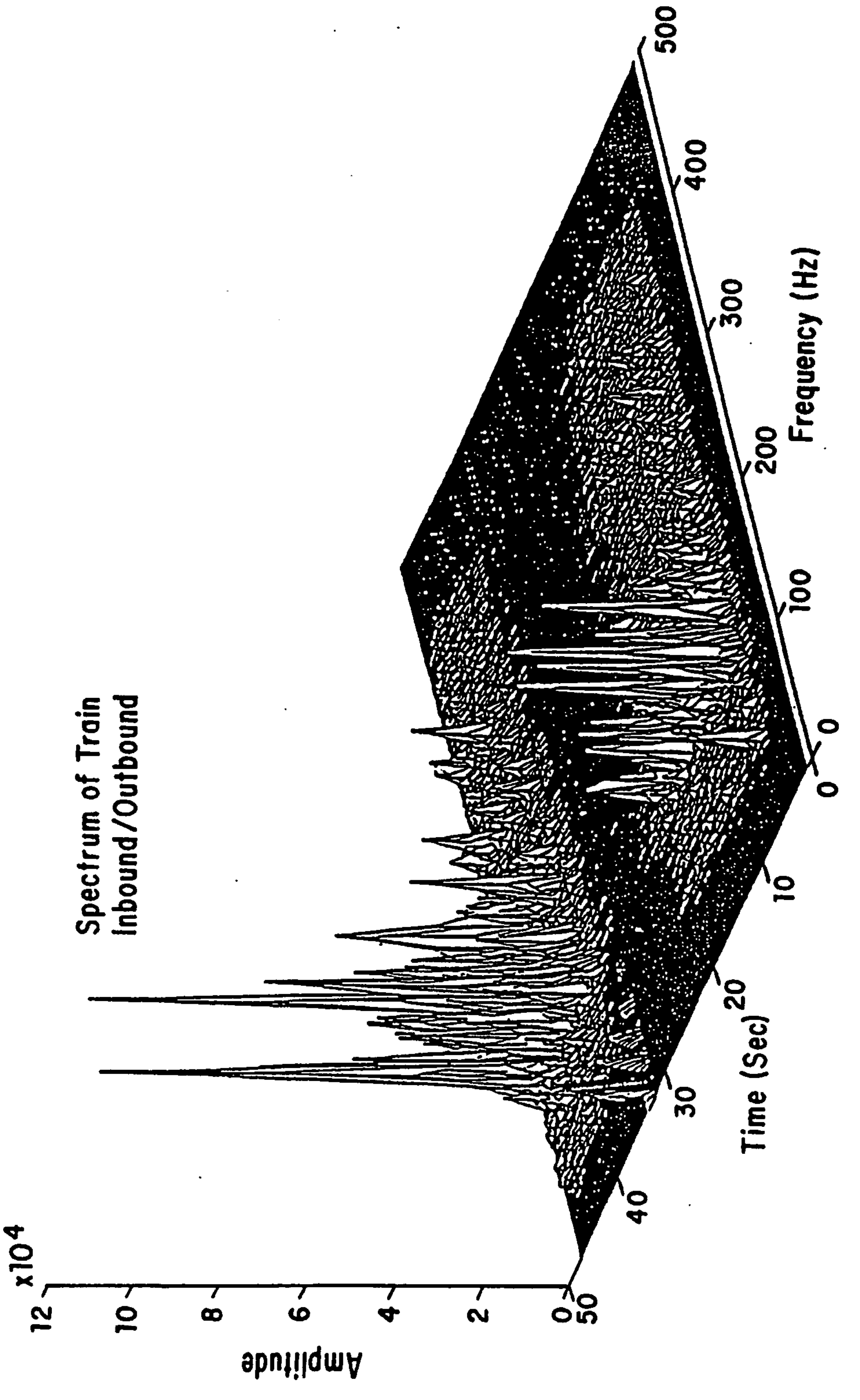

Spectrum of Train
Inbound/Outbound
Amplitude
x10^4
12
10
8
6
4
2
0
50
40
30
20
10
0
Time (Sec)
0
100
200
300
400
500
Frequency (Hz)

FIG. 13

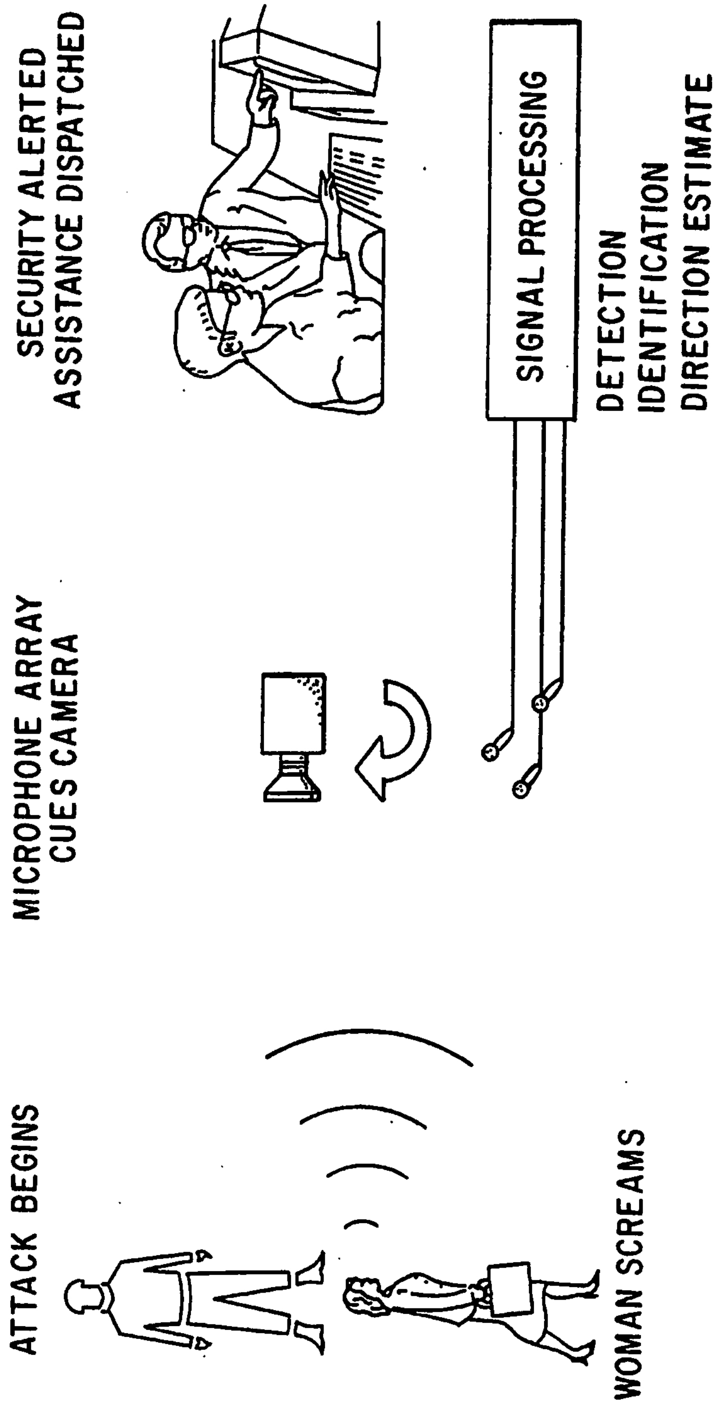

ATTACK BEGINS
WOMAN SCREAMS
MICROPHONE ARRAY
CUES CAMERA
SECURITY ALERTED
ASSISTANCE DISPATCHED
SIGNAL PROCESSING
DETECTION
IDENTIFICATION
DIRECTION ESTIMATE

FIG. 14a

Event	Classification	Action Pointer
Woman Scream	01	0110
Train Inbound	10	1001
Pistol	01	0110
*	*	*
*	*	*
*	*	*
*	*	*
*	*	*

FIG. 14b

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/10681

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :H04N 7/18
US CL :348/143
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 340/507, 521, 522, 540, 541, 544, 550, 905, 907, 937; 348/143

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 5,382,943 (TANAKA) 17 JANUARY 1995, FIGURE 4.	1-26
Y	US, A, 4,806,931 (NELSON) 21 FEBRUARY 1989, FIGURE 1.	1-26
A	US, A, 4,951,147 (AKNAR ET AL) 21 AUGUST 1990.	1-26
A	US, A, 5,428,345 (BRUNO) 27 JUNE 1995.	1-26
A	US, A, 4,857,912 (EVERETT, JR. ET AL) 15 AUGUST 1989.	1-26
A	US, A, 3,661,224 (ALLEN ET AL) 09 MAY 1972.	1-26
A	US, A, 5,414,409 (VOOSEN ET AL) 09 MAY 1995.	1-26

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 JANUARY 1996	15 FEB 1996

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	RICHARD LEE Joni Hill
Facsimile No. (703) 305-3230	Telephone No. (703) 308-6612

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/10681

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 5,376,919 (RICKMAN) 27 DECEMBER 1994.	1-26
A	US, A, 4,935,952 (DUTRA) 19 JUNE 1990.	1-26
A	US, A, 5,416,711 (GRAN ET AL) 16 MAY 1995.	1-26
A	US, A, 4,060,803 (ASHWORTH, JR.) 29 NOVEMBER 1977.	1-26
A	US, A, 5,323,141 (PETEK) 21 JUNE 1994.	1-26
A	US, A, 4,668,941 (DAVENPORT ET AL) 26 MAY 1987.	1-26
A	US, A, 4,241,335 (BARNES) 23 DECEMBER 1980.	1-26
A	US, A, 4,920,332 (PHILIPPE) 24 APRIL 1990.	1-26
A	US, A, 5,164,703 (RICKMAN) 17 NOVEMBER 1992.	1-26